NETWORK ANALYSIS OF STOCK MARKETS

A Dissertation
Presented to
The Academic Faculty

By

Fengrong Wei

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Scheller College of Business

Georgia Institute of Technology

August 2020

NETWORK ANALYSIS OF STOCK MARKETS

Approved by:

Dr. Cheol Eun, Advisor
Scheller College of Business
Georgia Institute of Technology

Dr. Suzanne Lee
Scheller College of Business
Georgia Institute of Technology

Dr. Alex Hsu
Scheller College of Business
Georgia Institute of Technology

Dr. Soohun Kim
Scheller College of Business
Georgia Institute of Technology

Dr. Vikas Agarwal
J.Mack Robinson College of Business
Georgia State University

Date Approved: April 28, 2020

To my loving family, who guided me to where I am today.

ACKNOWLEDGEMENTS

First, I would like to express my special appreciation and thanks to my advisor Professor Eun for the continuous support of my Ph.D. study. He have been a tremendous mentor for me. I would like to thank you for encouraging my research. Your advice on both research as well as on my career have been invaluable. His guidance helped me in all the time of my research and writing of this thesis.

Besides my advisor, I would also like to thank the rest of my committee members, professor Suzzane Lee, professor Alex Hsu, professor Soohun Kim, and professor Vikas Agarwal for serving as my committee members. I also want to thank you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you.

And I would also like to thank all the finance faculty and students. Over the years, they have been very helpful and supportive to me. Words can not express my gratitude.

Last but not the least, I would like to thank my family. Words can not express how grateful I am to my mother and father for all of the sacrifices that you've made on my behalf. Your prayer for me was what sustained me thus far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This dissertation consists of three essays on the application of network methods in finance study at the country-level, firm-level, and fund-level. In the first essay, we extend the analysis of globalization from the market factor to the rest of Fama-French factors and the Carhart momentum factor. The findings show that most of the sample local factors are significantly globalized, with the degree of globalization varying substantially across factors. Specifically, the market factor is the most globalized factor on average, followed by the momentum, size, value, profitability, and investment factors. In addition, we show that the impact of financial globalization has been imputed in the local factors, which explains the intriguing finding of integrated international asset pricing. That is, the local Fama-French factors outperform the global counterparts in pricing stocks, seemingly suggesting that stocks are priced as if financial markets were segmented despite the evident globalization. Our results indicate that this puzzle is attributable to the globalization of local factors.

In the second essay, we propose a system-wide approach to the study of the firm-specific connections, which capture the distinct relatedness between firms through unique features, conditional on the U.S. market. The proposed approach provides a new system-wide and factor-free measurement of market integration. We find that the degree of the firm-specific connections has decreased over time, and industry and style attributes significantly positively affect these connections. By applying these connections, investors can consistently gain through holding relatively few stocks randomly chosen across communities clustered based on these connections. Moreover, this consistency of gains has increased substantially over time, pointing to the importance of considering the firm-specific connections in risk diversification.

In the third essay, we use holding-linked network of mutual funds, measured by the similarity between funds' portfolios, to examine the network predictability of fund performance and flows. Using the new network method, we find evidence of significant pre-

dictability between funds with similar holdings. The predictability persists three to six months for alternative performance measures and at least twelve months for fund flows. In addition, a long-short strategy based on these holding links yields a significant annual alpha of about 4.5%. These findings reflect the similar underlying drivers of funds' portfolio holdings and show the persistent prediction of fund performance and flows by the holding linked network.

CHAPTER 1

GLOBALIZATION OF LOCAL FACTORS:IMPLICATIONS FOR THE ASSET PRICING

1.1 Introduction

A vast literature documents that the stock markets of most countries have become increasingly integrated.[1] A majority of prior work on integration focuses on local market factor. To extend the literature, we delve into other local factors, which may account for the "puzzle" residing in the intersection between the literatures on globalization and on integrated international asset pricing. Specifically, past research on integrated international asset pricing shows that Fama *et al.* (FF) factors are local (i.e., country-specific), rather than global, in that the local version of the FF factor model better explains the time-series variations in international stock returns than the global version in terms of pricing accuracy and explanatory power(e.g., [4], [5], [6]).[2,3,4] In line with this finding, [4] suggests that the estimation of the corporate cost of capital and the evaluation of the investment portfolio performance should be conducted based on the local factor model, not the global factor model. This finding seems to suggest that stocks are priced locally as if stock markets were segmented. While the finding is empirically robust, it is puzzling in light of the apparent globalization of the financial markets that has been observed in recent decades.

[1] For example, [1], [2].

[2] [4] shows that country-specific FF three-factor model outperforms the global version in explaining stock returns and portfolios in four countries, the U.S., the U.K., Canada, and Japan.

[3] [5] show that the local and international factor models typically outperform the global versions of the factor models in explaining the variation in local stock returns in most countries. The authors further show that the international factor model that includes the market, momentum, and cash flow-to-price factor mimicking portfolios provides the lowest rejection rate and pricing error, and ranks near the top in terms of model explanatory power among different versions of the multifactor models.

[4] [6] examine the performance of different versions of the Fama-French-Carhart four-factor model in pricing regional stock portfolios. We consider the Fama-French-Carhart four factor model in our main analysis and extend the analysis using the [7] five-factor model as a robustness check.

In this paper, we posit the conjecture that this puzzle arises from the globalization of the local FF factors and proceed to test this conjecture. Specifically, we first acknowledge that the Fama-French factors are, in fact, long-short portfolios comprised of stocks and thus are assets by themselves. As such, the pricing effect of global integration, in so far as globalization has occurred, should have been imputed in these factors. To the extent that the local FF factors became globalized, these factors were rendered to partly proxy the global factors. In this paper, we thus (i) test if the local FF factors are indeed globalized and (ii) study the implications of the globalized local factors for asset pricing.

In testing the globalization of local factors, we follow [2] who proposed the 'R-square method' for measuring global financial integration. Specifically, we estimate the degree of globalization of the local factors based on the adjusted R-squares from regressions of the local factors on the global factors. However, unlike Pukthuanthong and Roll who used as the global factors the principle components extracted from the stock market indices, we employ the pre-identified Fama-French-Carhart (FFC) six global factors [3], [8], [7] in our analysis. Thus, we estimate the R-square, which is our measure of global integration, by regressing each local factor in each of our sample countries on the global versions of the FFC factors, i.e., the market, size, value, momentum, profitability, and investment factors. A higher (adjusted) R-square thus obtained is interpreted as indicating a greater degree of globalization for the local factor during the regression period. In studying the implications of the globalized local factors for asset pricing, we focus on the relative performances of the three versions of the FFC factor model — the local, the global, and the pure (orthogonalized) local factor models.[5] The pure local factors are the residuals from the regressions of the local factors on the global factors. For our main analysis, the global factors are constructed by value-weighting the local factors of our nine sample countries.

We conduct our empirical analysis mainly using four major stock markets with the most

[5] We also examine the performance of the international factor model through incorporating both the foreign and local factors as [4] did. The results from the international factor model are very similar to those from the local factor model, as [4] found earlier. Detailed results from the international factor model are discussed in Section 5 when we revisit Griffin's study.

individual stocks, i.e., Canada, Japan, the United Kingdom, and the United States. These are the same four markets that [4] studied. During our whole sample period of 1990–2017, these four markets have sufficient individual stocks to allow us to form alternative sets of 5×5 portfolios sorted on (i) size and book-to-market, (ii) size and momentum, and (iii) size and R-square. Here, R-square captures the degree of global integration. These portfolios enable us to conduct meaningful cross-sectional analyses. But whenever deemed desirable, we complement these four markets with five additional markets — Australia, France, Germany, Hong Kong, and Switzerland — in order to establish the robustness of our empirical findings.

The key findings of our paper can be summarized as follows. First, the time-series mean of (adjusted) R^2, which is our measure of global integration, is significant for most of our sample local factors (i.e., 53 out of the total of 54 local factors), at the conventional significance levels, over the whole sample period of 1990–2017. That is, few of the local factors in our sample are "purely local" — most are globalized, albeit to varying degrees. Moreover, the degree of globalization of the local factors measured by the mean R^2 varies a great deal across the sample countries and factor types. Specifically, the mean R^2 ranges from 0.527 (Japan) to 0.879 (U.S.) for the local market factor, with an (cross-country) average of 0.665, while it ranges from 0.091 (Australia) to 0.702 (the U.S.) for the local size factor, with an average of 0.322, and ranges from 0.030 (Hong Kong) to 0.871 (the U.S.) for the local value factor, with an average of 0.285. On the other hand, the mean R^2 for the local momentum factor ranges from 0.176 (Hong Kong) to 0.876 (the U.S.) with an average of 0.454, while for the local profitability factor it ranges from 0.053 (Hong Kong) to 0.807 (the U.S.) with an average of 0.273, for the local investment factor it ranges from 0.058 (Hong Kong) to 0.810 (the U.S.), with an average of 0.268. Thus, on average, the market factor is the most globalized, followed by the momentum, size, value, profitability, and investment factors. Our sub-sample period results indicate that the mean R^2 tends to be higher in the later sub-sample period, April 2004– December 2017, than in the earlier sub-sample period, July 1990– March 2002, particularly for the market and momentum factors. It is also noted that the number of local factors with the significant R^2 rose from 48 (out of 54 local factors) in the early sub-sample period to 52 in the later sub-sample period; thus most local factors in our sample became globalized in the recent sub-sample period. Our findings pertaining to the globalization of the local factors are largely robust to alternative versions of the global factors constructed by (i) pooling all sample stocks across the nine sample countries, (ii) equal-weighting local factors of our sample countries, and (iii) applying the principle component analysis as Pukthuanthong and Roll (2009) did. Our findings indicate that most local factors in our sample countries are significantly globalized to a certain extent and, as a result, these local factors may partly proxy the effect of the global factors on asset pricing.

Second, using individual stock returns from our nine sample countries, we first compute the differences in the time-series means of the cross-sectional average pricing errors and explanatory powers between the local and global factor models over five-year rolling windows and then estimate the time trends in these differences. We find that the local factor model has a smaller mean pricing error than the global factor model for all sample countries and a greater mean R^2 than the global factor model for all sample countries, which are broadly consistent with the findings of [4]. Furthermore, we find "convergence" in both the pricing errors and explanatory powers between the local and global factor models over time in the majority of our sample countries, which again is indicative of the globalization of the local factors previously mentioned. Then we compute and compare the pricing performances of the three factor models — the local, global, and pure local models. We find that the absolute alpha is 1.73% for the local factor model, 1.90% for the global factor model, and 1.99% for the pure local model, on average (across countries) over our whole sample period. In terms of pricing accuracy, the local factor model performs the best and the pure local factor model the worst. This implies that the strong pricing performance of the local factor model can be attributable to the globalization of the local factors. The R^2, on the other hand, is 0.25 for the local factor model, 0.16 for the global factor model, and 0.08 for the pure local model, on average. Thus, the local factor model has the most explanatory power while the pure local factor model has the least, which is consistent with the results for the pricing accuracy.

Third, to further compare the performances of the alternative factor models, we form two sets of 25 (5×5) portfolios sorted on the size-BE/ME and size-MOM dimensions, respectively, and estimate the pricing errors and R^2s for the four countries (Canada, Japan, the U.K., and the U.S.) over the period from July 1990 to December 2017. Our results from the portfolios are generally consistent with those from the individual stocks. The pricing error is the lowest for the local factor model and the highest for the pure local model, while the R^2 is the highest for the local model and the lowest for the pure local model, for both sets of portfolios. We then expand our sample to nine countries and repeat our analysis using the two sets of nine (3×3) portfolios over the period from January 1996 to December 2017. Similar results are obtained.

Fourth, in order to test if the relative performance of the local vs. global factor models may differ across stocks with different degrees of globalization (measured by R^2), we form 25 size-R^2 sorted portfolios in each of the four major countries and estimate the pricing accuracy and explanatory power. Our results indicate that the difference in the pricing error and R^2 between the local and global factor models tends to decline "cross-sectionally" from portfolio 1 (containing small and segmented stocks) to portfolio 25 (containing large and integrated stocks). Our results also indicate that both the local and global factor models tend to perform poorly in terms of pricing accuracy for small and segmented stocks in most countries. These stocks may possibly be segmented not only globally but also domestically, making idiosyncratic risk a possible driver of its pricing.

Fifth, in an attempt to closely replicate the study of [4] and to reinterpret his findings in light of the globalization of the local FF factors, we estimate the pricing accuracy and explanatory power of the three competing models he considered, i.e., the domestic (local) factor model, the world (global) factor model, and the international (local and foreign) factor model, plus the pure local factor model. Each of these models is based on the Fama-French three factors. We estimate four versions of the FF three-factor models over the same sample period from January 1981 to December 1995 that was used in the Griffin study. The results indicate that the local FF factor model has a smaller pricing error and a greater explanatory power on average than the global FF factor model, and that adding foreign factors does not improve the performance of the local factor model. Thus, we successfully reproduce the key findings of [4]. Furthermore, we find that the pure local factor model performs the worst among the four factor models in terms of pricing accuracy and explanatory power. This implies that the outperformance of the local (domestic) version of FF three-factor model over the global (world) version documented by Griffin is attributable to the globalization of the local FF factors; it may not be reflective of the segmentation of international stock markets. The outperformance of the global FF factor model over the local FF factor model, net of the globalization effect, agrees well with the evident globalization of financial markets that has occurred over recent decades.

Overall, our findings indicate that the local factor model performs the best in terms of pricing accuracy and explanatory power, followed by the global factor model and the pure local factor model. The superior performance of the local factor model arises from the dual role that local factors have, namely, the pure local factors and proxies for the partial globalization effects. In fully integrated financial markets, stocks everywhere will be priced by a common set of global factors, as pointed out by [6], but in segmented markets, stocks will be priced by country-specific factors only. In partially integrated capital markets, however, the local FF factors might have the dual role, as mentioned above. Thus, our findings are indicative of the partially integrated nature of world financial markets.

Our findings also indicate that the relative performances among the competing factor models tend to vary across portfolios with different attributes in terms of the size, book-to-market ratio, degree of globalization, and country membership. In addition, our variance decomposition analysis of portfolio returns suggests that stocks may be priced simultaneously by the global and (pure) local factors (and also possibly idiosyncratic risks), with the relative importance among them varying across stocks/portfolios with different attributes. For practical applications of asset pricing models for the purpose of the cost of capital estimation and portfolio performance evaluation, it thus would be important to apply the pricing model that can consider realistic capital market regimes, which is likely to be that of partial integration allowing for the heterogeneous degrees of integration within and across countries.

The remainder of the paper is organized as follows. Section 1.2 describes the data and the construction of the local factors. Section 1.3 examines the globalization of the local factors. Section 1.4 evaluates and compares the performances of the local, global, and pure local factor models using individual stocks and portfolios. Section 1.5 revisits the key findings of Griffin (2002). Section 1.6 concludes the paper.

1.2 Data and Local Factor Construction

In this section, we introduce the data used in this study and then describe the construction method of the local factors.

1.2.1 Data

Our sample consists of publicly traded firms from nine developed countries: Australia, Canada, France, Germany, Hong Kong, Japan, Switzerland, the U.K., and the U.S. Our sample period is from July 1990 to December 2017. The sample countries are comprised of four countries from Asia/Pacific, three countries from Europe, and two countries from North America.

Data for the U.S. are obtained from the Center for Research in Security Prices (CRSP) and Compustat North America databases. Data for Canada are obtained from the Compustat North America database. Data for the other seven countries are obtained from the

Compustat Global database. For the U.S., firms whose common stocks are, or were, traded on the New York Stock Exchange (NYSE), American Stock Exchange, and NASDAQ are selected. For the other countries, firms whose common stocks are traded on each country's exchanges with the largest, and second-largest for some countries, number of traded stocks are selected. The exchanges selected are the ASX All Markets (EXCHGCD=106)[6] for Australia, the Toronto Stock Exchange (7) and TSX Venture Exchange (9) for Canada, the NYSE Euronext Paris (286) and Paris (231) for France, the Deutsche Boerse AG (154) and XETRA (171) for Germany, the Hong Kong Exchanges and Clearing Ltd (170) for Hong Kong, the Tokyo Stock Exchange (264) and Osaka Securities Exchange (227) for Japan, the Swiss Exchange (151) and Zurich (280) for Switzerland, and the London Stock Exchange for the U.K.

We restrict our analysis to the common stocks of firms trading in the home country of the firms. For the U.S. data, we restrict our sample to the common stocks of the firms that are incorporated in the U.S. and whose reporting currency is the U.S. dollar. For the non-U.S. data, we apply the screening and cleaning procedures suggested by [9], [10], [11]. Specifically, first, in order to exclude the non-common equity Compustat securities, we select the Compustat Issue Type TPCI='0'. We next apply the extensive generic and country-specific filter rules of [10] to the Compustat Issue Description DSCI. In addition, we exclude securities whose DSCI contain the "%" symbol ([12]) or 'INC.FD' ([13]). We then exclude investment funds and trusts by applying the procedure proposed by [12] to the Compustat Industry Codes GIND and GSUBIND variables and Business Description BUSDESC variable.

For the U.S., monthly returns represent delisting-adjusted returns, i.e., return plus delisting return from CRSP. For the other countries, we build total return indices in U.S. dollars using the Compustat price variable PRCCD, adjustment factor ADJEXDI, quotation unit QUNIT, total return factor TRFD, and exchange rate EXRATD. Specifically, we first apply

[6]"EXCHGCD" represents the exchange code in the Compustat database.

the decimal error cleaning procedure of [12] to PRCCD, ADJEXDI, QUNIT, and TRFD. Then, after transforming EXRATD to exchange rates representing U.S. dollar per one unit of other currency, we build total return indices in U.S. dollars using $(PRCCD/QUNIT)*(TRFD/AJEXDI)*EXRATD$.[7] Using the total return indices at two successive month ends, we construct monthly returns in U.S. dollars. For non-U.S. stocks that are indicated to have been delisted due to bankruptcy (DLRSNI='02') or liquidation (DLRSNI='03') from the Compustat security descriptor data, we use delisting-adjusted returns in the delisting month (DLDTEI), i.e., return plus -30% following [14]. Finally, for all sample countries including the U.S., monthly returns below -100% and above 300% are set to missing to filter out errors and reduce the influence of extreme returns (e.g., [15]).

Throughout this paper, returns are in U.S. dollars, and excess returns are returns in excess of the one-month U.S. Treasury bill rate obtained from the CRSP (e.g. [6]; [16]).

1.2.2 Construction of Local Factors

The local market factor (denoted by MKT) of each country is constructed through subtracting the one-month U.S. Treasury bill rate from the local market return. For each country, the local market return of a month represents the market-capitalization weighted average of monthly stock returns with non-missing monthly returns for the month and positive market capitalizations at the prior month end.

The local size and value factors (denoted by SMB and HML, respectively) of each country are constructed following the standard method of [17, 3, 18, 7]. At the end of June of each year t, the stocks within the country are sorted according to their current size, proxied by the market capitalization, and the book-equity-to-market-equity ratio (BE/ME) of year $t-1$. For the U.S. firms, BE/ME is based on the book value of common equity

[7] The market capitalization of a non-U.S. stock is calculated as (PRCCD/QUNIT)*CSHOC*EXRATD, where CSHOC represents the Compustat shares outstanding variable. For Canada, CSHOC is missing prior to April 1998 in the Compustat North America daily data files. Thus, for the pre-1998 period, we use the CSHOQ variable instead which is available at quarter ends from Compustat North America Fundamental Quarterly data files.

at the fiscal year-end of year $t-1$ and the market capitalization at the end of December of year $t-1$. The BE value is calculated following [19].[8] We require positive size and BE/ME values. Then, we construct six value-weighted portfolios from independent 2-by-3 size-BE/ME sorts. We classify firms into two groups using the size breakpoint of the 50th percentile—small (S) and big (B)—and into three groups using the BE/ME breakpoints of the 30th and 70th percentiles—high (H), median (M), and low (L).[9] The intersections of the independent 2-by-3 size-BE/ME sorts produce six value-weighted portfolios: SH, SM, SL, BH, BM, and BL. The portfolios are held over the one-year period from July of year t to June of year $t+1$. The SMB factor (small minus big) is computed using SMB = (SH+SM+SL)/3 - (BH+BM+BL)/3. The HML factor (high minus low) is computed using HML = (BH+SH)/2 - (BL+SL)/2.

The local momentum factor (denoted by MOM) of each country is constructed following the method suggested in the website of Kenneth French.[10] At the beginning of each month m, the stocks within the country are sorted according to their current size and prior (2-12) return, i.e., the return from the end of month $m-13$ to the end of month $m-2$. We require a positive size and valid prior (2-12) return values. We classify firms into two groups using the size breakpoint of the 50th percentile—small (S) and big (B)—and into three groups using the prior (2-12) return breakpoints of the 30th and 70th percentiles—high (U), median (M), and low (D). The intersections of the independent 2-by-3 size-prior return sorts produce six value-weighted portfolios: SU, SM, SD, BU, BM, and BD. The portfolios are updated monthly. The MOM factor is defined as MOM = (BU+SU)/2 - (BD+SD)/2.

We construct the local profitability and investment factors of each country by following the method of [7]. At the end of June of each year t, we construct six value-weighted port-

[8]For the non-U.S. firms, as for preferred stocks, redemption and par values are available, but liquidation values are not available.

[9]For the U.S., only NYSE-listed stocks are used to compute the breakpoints. This rule applies equally to the construction of all other U.S. factors.

[10]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html.

folios from independent 2 by 3 size-profitability sorts and size-investment sorts following the same method as used for constructing size and value factors above. The profitability of a firm is calculated as the annual sales revenues (SALE) minus cost of goods sold (COGS), selling, general, and administrative expenses (XSGA), and interest expense (XINT) divided by book equity (BE) for the fiscal year-end of year $t-1$. For the computation of the firm profitability, we require positive book equity (BE) data, non-missing revenues data, and non-missing data for at least one of the following: COGS, XSGA, and XINT. The investment of a firm is calculated as the change of total assets (AT) from the fiscal year-end of year $t-2$ to the fiscal year-end of year $t-1$ divided by $t-2$ total assets. For the computation of the firm investment, we require total assets data for year $t-2$ and $t-1$. The construction of the profitability and investment factors are exactly the same as that of the value factor, and they are denoted by RMW (robust minus weak) and CMA (conservative minus aggressive) following [7].

Table 1.1 presents the summary statistics of the monthly returns in U.S. dollar terms for the local MKT, SMB, HML, MOM, RMW, and CMA factors from each of the nine sample countries and also the value-weighted global MKT, SMB, HML, MOM, RMW, and CMA factors. All reported statistics are based on the monthly time series. As can be seen from the table, the (time-series) mean of the local MKT factor ranges widely from 0.11% for Japan to 0.87% for Hong Kong. Similarly, the mean return on the local SMB factor varies widely across countries, ranging from -0.14% for Germany to 0.46% for Australia. Moreover, the mean return on the local SMB is positive for seven sample countries (i.e., Australia, Canada, France, Hong Kong, Japan, the U.K., and the U.S.), but negative for the remaining two countries (i.e., Germany and Switzerland). Apparently, there was a "large-cap premium" in each of the latter two countries during our sample period, which is an intriguing phenomenon.[11] The mean return of the local HML factor, on the other hand,

[11] According to [20], the increased presence of institutional investors in recent decades and their increased demand for stocks of large companies can explain part of the disappearance of the small-company stock premium and the appearance of the large-company stock premium.

exhibits a modest variation across countries, ranging from 0.04% for Switzerland to 0.53% for Australia and Japan. Among the six local factors, the mean return for the MOM factor exhibits the largest variation across countries, ranging from 0.11% for Japan to 1.74% for Australia; as is well known, the MOM factor is insignificant in Japan. The mean return on the local RMW factor ranges from 0.04% for Japan to 0.64% for Canada, and the mean return on the local CMA factor ranges from -0.20% for Hong Kong to 0.56% for Canada.

[Table 1.1 inserted here]

Each of the global factors presented in Table 1.1 is constructed as the value-weighted average of the corresponding country-specific, local factor, with the weight computed as the fraction of the country's market capitalization in the aggregate market capitalization of the sample countries in the previous period. The (time-series) mean return of the global factor is the highest for the MOM factor (0.57%), followed by the MKT (0.49%), HML (0.30%), RMW (0.28%), CMA (0.18%), and SMB (0.12%) factors, while the time-series standard deviation is the highest for the MKT factor (4.22%), followed by the MOM (3.87%), HML (2.30%), SMB (2.27%), RMW (1.89%), and CMA (1.38%) factors. The Sharpe ratio is the highest for the MOM and RMW factors (0.15), followed by the HML and CMA factors (0.13), MKT (0.11), and SMB (0.05) factors.

Table 1.1 also reports the average number and size of the sample firms that have valid returns and capitalization data. As shown in the table, both the average number and size of the sample firms vary a great deal across countries. In particular, the U.S. represents the largest number of sample firms (4,590), followed by Japan (2,693), the U.K. (985), and Canada (692), with Switzerland hosting the smallest number of firms (163). However, the average firm size is the largest for Switzerland ($3,265.62 m) and the smallest for Australia ($1063.87 m).

1.3 Globalization of Local Factors

In this section, we describe how we measure the degree of global integration and then we estimate the global integration of the local factors for each of the nine sample countries, focusing on these two aspects, i.e., the degree of globalization and the time-trends in globalization.

1.3.1 Measuring the Globalization of Local Factors

Following [2] who proposed the "R-square method" for measuring global financial integration, we estimate the degrees of globalization of local factors based on the adjusted R-squares from the regressions of the local factors on the global factors.[12,13] However, unlike [2] who used as the global factors the principle components extracted from the stock market indices, we employ the pre-identified global versions of the Fama, French, and Carhart (hereafter, FFC) six factors in our analysis. Specifically, we estimate the R^2 through regressing each local factor from each of the sample countries on the global versions of the FFC factors comprising the MKT, SMB, HML, MOM, RMW, and CMA factors. A higher R^2 thus obtained is interpreted as indicating a greater degree of globalization for the local factor during the regression period.

Beginning in July 1990, the local factor regressions are estimated each month over

[12] [2] show a fundamental flaw in using the cross-country correlations of the stock index returns as a measure of integration and propose a new measure of integration based on the explanatory power of a multi-factor model, i.e., R-squares. They show that the perfectly integrated markets can exhibit a weak correlation whenever there are multiple global sources of country index return volatility and the countries do not share the same sensitivities to all of them.

[13] In this paper, we use the adjusted R-squares from the regressions for all analyses and use R^2s in writing, hereafter, for simplicity.

60-month rolling windows as follows:

$$LMKT_{i,t} = \alpha_{i,M} + \beta^{GM}_{i,M} GMKT_t + \beta^{GS}_{i,M} GSMB_t + \beta^{GH}_{i,M} GHML_t \tag{1a}$$
$$+ \beta^{GO}_{i,M} GMOM_t + \beta^{GR}_{i,M} GRMW_t + \beta^{GC}_{i,M} GCMA_t + \epsilon_{i,M,t},$$
$$LSMB_{i,t} = \alpha_{i,S} + \beta^{GM}_{iS} GMKT_t + \beta^{GS}_{i,S} GSMB_t + \beta^{GH}_{i,S} GHML_t \tag{1b}$$
$$+ \beta^{GO}_{i,S} GMOM_t + \beta^{GR}_{i,S} GRMW_t + \beta^{GC}_{i,S} GCMA_t + \epsilon_{i,S,t},$$
$$LHML_{i,t} = \alpha_{i,H} + \beta^{GM}_{i,H} GMKT_t + \beta^{GS}_{i,H} GSMB_t + \beta^{GH}_{i,H} GHML_t \tag{1c}$$
$$+ \beta^{GO}_{i,H} GMOM_t + +\beta^{GR}_{i,H} GRMW_t + \beta^{GC}_{i,H} GCMA_t + \epsilon_{i,H,t},$$
$$LMOM_{i,t} = \alpha_{i,O} + \beta^{GM}_{i,O} GMKT_t + \beta^{GS}_{i,O} GSMB_t + \beta^{GH}_{i,O} GHML_t \tag{1d}$$
$$+ \beta^{GO}_{i,O} GMOM_t + +\beta^{GR}_{i,O} GRMW_t + \beta^{GC}_{i,O} GCMA_t + \epsilon_{i,O,t},$$
$$LRMW_{i,t} = \alpha_{i,R} + \beta^{GM}_{i,R} GMKT_t + \beta^{GS}_{i,R} GSMB_t + \beta^{GH}_{i,R} GHML_t \tag{1e}$$
$$+ \beta^{GO}_{i,R} GMOM_t + +\beta^{GR}_{i,R} GRMW_t + \beta^{GC}_{i,R} GCMA_t + \epsilon_{i,R,t},$$
$$LCMA_{i,t} = \alpha_{i,C} + \beta^{GM}_{i,C} GMKT_t + \beta^{GS}_{i,C} GSMB_t + \beta^{GH}_{i,C} GHML_t \tag{1f}$$
$$+ \beta^{GO}_{i,C} GMOM_t + +\beta^{GR}_{i,C} GRMW_t + \beta^{GC}_{i,C} GCMA_t + \epsilon_{i,C,t},$$

where $LMKT_{i,t}, LSMB_{i,t}, LHML_{i,t}$, $LMOM_{i,t}$, $LRMW_{i,t}$, and $LCMA_{i,t}$ are the local MKT, SMB, HML, MOM, RMW, and CMA factors of the i-th country in month t, respectively; $GMKT_t, GSMB_t, GHML_t$, $GMOM_t$, $GRMW_t$, and $GCMA_t$ are the corresponding global factors in month t; the βs with superscripts "GM", "GS", "GH", "GO","GR", and "GC" represent the exposure coefficients for the global MKT, SMB, HML, MOM, RMW, and CMA factors, respectively; and the subscripts for α, β, and ϵ – "M", "S", "H", "O", "R", and "C" – refer to the dependent variables, i.e., local MKT, SMB, HML, MOM, RMW, and CMA factors, respectively. The last five-year window ends in December 2017. By restricting each window to five years of monthly observations, we allow for the possibility that the β exposures in equations (1a) to (1f) change over time. Then, the time-series average of R^2 from each local factor regression represents the degree of globalization of the local factor on average.

In order to formally test the significance of the globalization of each local factor, we apply an adjusted t-test. Because the R^2s are estimated over 60-month rolling windows moving one month forward at a time for each local factor, the estimated R^2s can be serially correlated. The effect of the serial correlation is, usually, to make the comparisons of the means liberal in the sense that significant differences are found more frequently than expected when there is no difference. Considering the serial correlation in the test of the mean, we adjust the t-statistics by Newey West method ([21], [22]). [14]

The time-series mean of the R^2s and the significance from the adjusted t-test aid us to learn the degrees of globalization for distinct local factors across different countries during the sample period. The existing literature has predominantly used local MKT factors to study the integration of capital markets (e.g., [1], [23], [2], [24]). The degrees of globalization for the local MKT factors are used to measure the degrees of capital market integration. In this paper, we examine the globalization of not only the local MKT factors but also the local SMB, HML, MOM, RMW, and CMA factors.

Next, in order to formally investigate the behavior of the globalization of each local factor over time, we investigate the time trend of the R^2s from each local factor regression during the whole sample period and compare the time-series mean of the R^2s from each local factor regression during two equally divided sub-sample periods. Considering the serial correlation of R^2s from each local factor regression during each study sample period, we again apply the adjusted t-test for the null hypothesis that the slope of the time trend equals zero and the null hypothesis the mean of the R^2s from each local factor regression during the recent sub-sample period is not higher than the mean of the R^2s during the early sub-sample period.

[14] In our main analysis, we adjust the t-statistics using 59 lags.

1.3.2 Value-weighted Local Factors as the Global Factors

To estimate equations (1a) to (1f), we first construct the global factors using the value-weighted averages of all sample local factors, following [4] and [6]. For example, the global MKT factor is formed using $GMKT_t = \sum_j \omega_{j,t-1} LMKT_{j,t}$, where $LMKT_{j,t}$ is the local MKT factor of the j-th country in month t, $\omega_{j,t-1}$ is the fraction of the dollar-denominated market capitalization of the j-th country in the total market capitalization of all nine sample countries in the previous month.[15] Similarly, the global SMB, HML, MOM, RMW, and CMA factors are constructed using $GSMB_t = \sum_j \omega_{j,t-1} LSMB_{j,t}$, $GHML_t = \sum_j \omega_{j,t-1} LHML_{j,t}$, $GMOM_t = \sum_j \omega_{j,t-1} LMOM_{j,t}$, $GRMW_t = \sum_j \omega_{j,t-1} LRMW_{j,t}$, and $GCMA_t = \sum_j \omega_{j,t-1} LCMA_{j,t}$, where $LSMB_{j,t}$, $LHML_{j,t}$, $LMOM_{j,t}$, $LRMW_{j,t}$, and $LCMA_{j,t}$ are the local SMB, HML, MOM, RMW, and CMA factors of the j-th country in month t, respectively. Alternative versions of the global factors are also used in the regressions in subsection 3.3, which enables the verification of the robustness of the globalization of the local factors through alternative ways of constructing the global factors.

The time-series R^2s from the five-year rolling window regressions in equations (1a) to (1f) for each local factor from each of the sample countries during the whole sample period are plotted in Figure 1.1. As can be seen from the figure, the local factors exhibit different dynamic patterns and degrees of global integration, which are measured by the R^2s.[16] In general, the local MKT and MOM factors exhibit greater degrees of global integration than the local SMB, HML, RMW, and CMA factors in most sample countries. Both the MKT and MOM factors exhibit upward trends in the majority of the sample countries in the early years, but exhibit downward trends in recent several years, especially the MKT factor. In contrast, the local SMB, HML, RMW, and CMA factors exhibit no discernable secular

[15] Alternatively, we employ the global factors used in [6] that are constructed from 23 countries for the analysis. The results are qualitatively similar to those from using the nine countries in our research. Detailed results are available upon request.

[16] As shown in Figure 1, the time-trends in the globalization for the local factors are time-varying, but not always linear. We thus compare the means of the R^2s between the two sub-sample periods in addition to the conventional linear time-trend analysis.

globalization trends in most countries during our sample period.

[Figure 1.1 inserted here]

In order to examine the globalization of the local factors in detail, we compute the time-series means of the (adjusted) R^2s from estimating the five-year rolling window regressions in equations (1a) to (1f) for the six local factors from each of the sample countries over the whole sample period and also over the two equally divided sub-sample periods of July 1990– March 2004 and April 2004– December 2017. The results are reported in Table 1.2 with the conventional linear time-trend analysis of the R^2s.

[Table 1.2 inserted here]

There are some noteworthy features in Table 1.2. First, the R^2s, which is our measure of globalization, are all positive and significant at the 10% level or better over the whole sample period of 1990–2017, with the sole exception of the local HML factor of Hong Kong. That is, 53 out of the total sample of 54 local factors are found to be significantly global. This result indicates that most local FFC factors are significantly globalized and, as a result, few local factors are "purely" local. In turn, this situation suggests that an observed local factor may have a dual role, reflecting both the pure local factor and the effect of globalization. This interpretation of the local factors seems to offer a clue to solving the integrated international asset pricing puzzle as [4] an example. We will revisit this possibility in the next section. Second, the degree of globalization as measured by R^2 varies a great deal across factors, countries, and time periods. We elaborate this point below.

Panel A of Table 1.2 illustrates that for each factor type, the degree of globalization (R^2) varies substantially across countries. In particular, the R^2 ranges from 0.527 (Japan) to 0.879 (U.S.) for the MKT factor, 0.091 (Australia) to 0.702 (U.S.) for the SMB factor, 0.030 (Hong Kong) to 0.871 (U.S.) for the HML factor, 0.176 (Hong Kong) to 0.876 (U.S.)

for the MOM factor, 0.053 (Hong Kong) to 0.807 (U.S.) for the RMW factor, and 0.058 (Hong Kong) to 0.810 (U.S.) for the CMA factor. On the other hand, the cross-country average R^2 is 0.665 for the MKT factors, 0.454 for the MOM factors, 0.322 for the SMB factors, 0.285 for the HML factors, 0.273 for the RMW factors, and 0.268 for the CMA factors. Thus, the MKT factors are the most globalized on average, followed by the MOM factors, the SMB factors, while the HML, RMW, and CMA factors are comparatively less globalized. In all sample countries, the MKT factor is substantially more globalized than the other factor types over the whole sample period. Among the sample countries, the United States is obviously the most integrated market across the different types of local factors, followed by the United Kingdom, whereas Hong Kong is the least integrated market, with the remaining countries falling in between. The finding that the Anglo-American markets are substantially more globally integrated than the Asian markets seems to be consistent with the general perception of these markets.

The time-series trend analysis of the R^2s in Panel A of Table 1.2 indicates that the degree of globalization increased over time for the MKT and MOM factors for most countries, while for the SMB, HML, RMW, and CMA factors, the degree of globalization demonstrated different trends. For example, there were no significant changes for the SMB factors in Austtralia, Canada, Germany, Hong Kong, Switzerland, the U.K., and the U.S., but increasing degrees of globalization in France, and decreasing degree of globalization in Japan.

Since the dynamic trends of the R^2s are not linear as illustrated in Figure 1.1, we compare the R^2s estimated from the two sub-sample periods and test the null hypothesis that the degree of globalization is greater in the recent sub-sample period than in the earlier sub-sample period for each local factor on average . The results are presented in Panel B of Table 1.2. As shown in the table, the degree of globalization increased over time for the MKT and MOM factors in all sample countries except the MKT factor of Japan, with these increases statistically significant in all sample countries except Japan. For the other local factors, there is no consistent trend across countries. Consistent with these results, the cross-country average R^2 increased from 0.522 during the first sub-sample period to 0.787 during the second sub-sample period for the MKT factor, which is an increase of about 51%, and from 0.329 to 0.568 for the MOM factor, which is an increase of about 73%, over the two sub-sample periods.

These two types of local factors, particularly the MOM factor, experienced very rapid globalization over time. Interestingly, the MOM factor exhibits low degrees of globalization during the first sub-sample period in most sample countries, which is consistent with [25], but shows much higher degrees of globalization during the second sub-sample period in all countries. In contrast, the cross-country average R^2 declined slightly from 0.349 to 0.327 for the SMB factor and 0.273 to 0.247 for the RMW factor over the same sub-sample periods.[17] At the country level, Australia experienced a significant increase in R^2 over time for five of the six local factors, significantly globalizing in most dimensions. In contrast, Japan experienced a significant decrease in the average R^2 in all factor types over time except the MOM factor, which implies that globalization stalled or diminished in Japan, which could reflect the country's so-called "lost decades" that largely coincide with our sample period.

Noteworthy from Table 1.2 is the finding that among all sample countries, the U.S. has the highest degree of globalization in each of the six categories of local factors over the whole sample period. This finding holds in each of the two sub-sample periods. This can result from the highly globalized nature of the U.S. market and/or the dominant market capitalization of the U.S. compared with other countries. Because the U.S. has much larger market capitalization compared with the other sample countries, the local factors of the U.S. would comprise large proportions of the global factors when the value-weighted local factors are used as the global factors. In order to mitigate the possibly disproportionate ef-

[17] The comparison of the time-series mean R^2s between the two sub-sample periods is based on one-sided tests. We test whether there is an upward pattern from the first to the second sub-sample period. We also examine the existence of a downward pattern. For example, it is found that the local SMB and CMA factors of the U.S. show significantly downward trends.

fect of the market capitalization on the global factors, we alternatively construct the global factors as the equal-weighted averages of the local factors of all sample countries and re-estimate the degrees of globalization. In order to conserve space, we tabulate the results in the Appendix (Panel A of Table A.2).

As can be seen from the Appendix, the key findings from using the equal-weighted global factors, which are directly relevant to our analysis, remain qualitatively similar to those from using the value-weighted global factors. In particular, all local factors are found to be significantly globalized over the whole sample period, but the degree of globalization varies substantially across factors and countries. Furthermore, the degree of globalization increases over time, particularly for the MKT and MOM factors, which is also found when the value-weighted global factors are used. It is noted, however, that a few interesting changes occur – when the equal-weight approach is used, for example, the degree of globalization increases for all sample countries, except for Japan and the U.S. which contribute the most to the global market capitalization in the value-weight case.

Given the regression structure, i.e., equations (1a) to (1f), that we use to estimate the R^2s, both the heterogeneous levels and trends of the globalization of the local factors reported in this section would be partly related to the beta exposures of the local factors to the global factors. Thus, it would be useful to examine the beta exposures. Table 1.3 presents the time-series averages of the beta exposures of each local factor to each of the six global factors over the whole sample period and the proportions of significant exposures over all five-year rolling windows at the 5% level. Same analysis of the beta exposures over the two sub-sample periods are conducted.[18]

[Table 1.3 inserted here]

Table 1.3 has some noteworthy features. First, each local factor has the highest beta exposure to its corresponding global factor. For example, the local MKT factor of Australia has a beta exposure of 1.10 to GMKT, but the exposure is only 0.04 to GSMB, 0.28 to

[18]Detailed proportions of the significant beta exposures are available upon request.

GHML, 0.05 to GMOM, -0.24 to GRMW, and -0.55 to GCMA, and the local SMB of the country has a beta exposure of 0.44 to GSMB, but only -0.03 to GMKT, 0.14 to GHML, -0.06 to GMOM, -0.23 to GRMW, and -0.20 to GCMA, etc. Furthermore, the beta exposures of local MKT, SMB, and MOM factors to the corresponding global factors are statistically significant at the 5% level at least 50% of the time over the estimation windows only except the SMB factor of Australia. Second, the "cross-exposures" of the local factors to the global factors of different types are relatively low and mostly insignificant over the whole sample period.

In addition, the comparison of the beta estimates between the two sub-sample periods shows the cross-exposure increased over time. The comparison also shows that the exposures of many local factors to the corresponding global factors became larger and more frequently significant in more recent years. These heterogeneous and changing beta exposures of the local factors to the global factors should have had a significant role in driving the levels and trends of the global integration of the local factors.[19]

1.3.3 Alternative Versions of the Global Factors

One potential concern in the preceding analysis is that the dependent variables (local factors) in regressions in equations (1a) to (1f) are included in the independent variables (global factors) due to the way that the global factors are constructed, regardless of whether the value- or equal-weighted local factors are used to construct the global factors. Another potential concern is that the differences across countries in terms of market capitalization, accounting conventions, etc., might influence our results. In order to verify the robustness of our results to the construction of the global factors, we employ two alternative methods to construct the global factors.

[19] From the definition of R^2, besides the βs in the regressions in equations (1a) to (1f), the variance of the local factors, the variances of the global factors, the covariances between the local factors and the global factors, and the covariances between different global factors could also affect the R^2 estimation. We verify these variances and covariances and find that they vary over time. For example, the global MKT and MOM factors demonstrate increasing correlations over our whole sample period, but the correlations become negative during the recent sub-sample period.

First, we construct the global factors directly from pooling stocks from all sample countries and forming portfolios.[20] Then, the global factors formed this way are used in the local factor regressions in equations (1a) to (1f). We obtain qualitatively similar results to those obtained from using the value-weighted global factors that are plotted in Figure 1.1 and also reported in Table 1.2. The detailed regression results are provided in the Appendix (see Panel B of Table A.2). As can be seen from the Appendix, the average R^2 is significant at the conventional levels for all local factors, except for the CMA local factor of Hong Kong and the RMW factor of Switzerland. That is, most local factors are significantly globalized, which is consistent with the previous results. In the first sub-sample period, the R^2s for all local MKT, SMB, MOM, and CMA factors are statistically significant, those for the local HML and RMW factors are mostly significant except the HML factor of Australia, Germany, and Hong Kong, the RMW factors of Hong Kong and Switzerland. In the second sub-sample period, however, the R^2s are significant for all local factors, except for the local SMB factor of Australia, the HML factor of Hong Kong, the RMW factor of France, Germany, and Switzerland, and the CMA factor of Germany.[21] A comparison between the sub-sample periods shows that both the local MKT and MOM factors increased sharply over time, while the local SMB, HML, RMW, and CMA factors do not have a consistent trend across countries, which is consistent with the previous results.

Next, we construct the global factors in equations (1a) to (1f) employing the method of the principal components used in [2]. For each type of local factors, we construct the principal components separately. Following [2], we first pool the local MKT factors from nine countries and use the 60-month returns of the local MKT factors from July 1990 to June 1995 to compute the weightings; then, the estimated weightings are applied to the returns of the local MKT factors in July 1995.[22] We repeat this process monthly and

[20] We construct the global factors from the formed portfolios using the same method as we used to construct the local factors in Section 1.2.

[21] We notice that a small value of R^2 does not necessarily mean that the degree of globalization is insignificant. For example, the degree of globalization for the local HML factor of Australia is 5.9% and significant at the 10% level.

[22] We follow the same method as in [2] to construct the principal components. They apply the previous

produce the out-of-sample principal components for the local MKT factors. As proxies for the global MKT factor, we retain the first two principal components from the local MKT factors as the global market factors, which explain 80.83% of the total volatility in the covariance matrix of the local MKT factors on average. We form the global SMB, HML, MOM, RMW, and CMA factors in the same way, with the first two principal components explaining 60.45%, 54.38%, 63.64%, 49.36%, and 66.38% of the total volatility in the covariance matrix of the local SMB, HML, MOM, RMW, and CMA factors on average, respectively.[23] In our application of the principal component method, twelve global factors are used in the regressions from July 1995 to December 2017.[24] In order to conserve space, the results are not tabulated in this paper.[25] Basically, the main findings reported previously remain robust to this alternative version of global factors. For example, the average R^2s for the local MKT factors range from 0.553 for Japan to 0.871 for France, and the average R^2s of the local SMB factors range from 0.072 for Japan to 0.643 for Hong Kong over the whole sample period.

Overall, our findings in this section indicate that most of the sample local factors are significantly globalized, with the degree of globalization varying substantially across factors, countries, and sample periods. These findings imply that the effect of globalization is likely to have been imputed in the local FFC factors that are constructed as long-short portfolios comprising stocks. In the next section, we investigate the implications of the globalized local factors for asset pricing.

---

one-year daily data to compute the weightings and apply the estimated weightings to the current year's local market indices to form the principal components.

[23]The first two principal components from each local factor explain the lower (higher) proportion of total volatility at earlier (later) five-year windows than the average proportion. Due to the five-year window regression, there are 60 monthly observations in the regression. If we extract more principal components, we will have a large number of global factors in the regressions. Moreover, more principal components will provide higher R^2s in the regressions and stronger results than the two-principal components.

[24]We also apply the method of principal components through pooling all local factors from all countries (54 factors) and retaining the first 22 principal components that explain at least 90% of the total volatility in the covariance matrix of all local factors in June 1995. Qualitatively similar results are obtained. Moreover, we repeat this analysis by excluding the local factor used as the dependent variable from the 54 local factors in order to extract the principal components. The degrees of globalization for the local factors (R^2s) are relatively smaller than R^2s estimated from using all 54 local factors to extract the principal components.

[25]The detailed results are available upon request.

1.4 Asset Pricing Tests

In this section, we comprehensively examine the ability of the local, global, and pure local factors to explain the stock returns, considering the globalized nature of the local factors, which explains the puzzle of international asset pricing. First, we construct the pure local factors through decomposing the local factors into globalized parts of the local factors and the pure local factors. We regress each local factor on the global factors over the whole sample period and take the residual from the regression as the pure local factor.[26] Then we evaluate the performance of the local, global, and pure local versions of the FFC six-factor model in explaining the individual stock and portfolio returns, and we investigate which factor model best explains the time-series and cross-sectional variations in the individual stock and portfolio returns. Comparison of the performances among the local, global, and pure local factor models also reveals the effects of globalization on the local factors.

We evaluate the performance of each factor model using the magnitude of the model pricing error (absolute alpha) and the explanatory power (R^2). A better factor model would have a lower model pricing error and a higher explanatory power. We also examine the time-trends in the differences between the performance of the local and global versions of the FFC six-factor model regressions in order to assess the joint globalization effects of all local factors. For example, if a time trend shows that the differences in the performances of the local and global factor models are reduced over time, it would imply that the ability of the local factors to explain the stock returns converges to that of the global factors over time. As an additional measurement of the model performance for portfolios, we also apply the Gibbons *et al.* (GRS) F-test statistics to test the hypothesis that the intercepts are jointly equal to zero across the test assets of interest.

In the following asset pricing tests, we employ the global factors that are formed by value-weighting the local factors of the sample countries. The local factors employed in

[26]We use the pure local factor as an analytical tool to study the impact of globalization on the local factors in this paper. We do not suggest that the pure local factors be operational factors in the asset pricing models.

the local version of the FFC model are also value-weighted local factors following [4].[27,28] For the j-th country local factor model, for example, the market capitalization weight $\omega_{j,t-1}$ is applied to each local factor in the country, i.e., the country local MKT factor is $\omega_{j,t-1} LMKT_{j,t}$. The same weight is applied to the remainder of the local factors of the country. This facilitates comparison with [4] in Section 5.

1.4.1 Individual Stock Returns

The dependent variables in the factor model regressions are the dollar-denominated stock returns in excess of the U.S. risk-free rate. Each month, beginning in July 1990, individual stock regressions are estimated over 271 rolling 60-month periods. In order to enter a regression, a stock must have all 60 monthly observations available during the five-year period. Because we use five-year rolling windows, the intercept and coefficients of all factors in a factor model are allowed to change over time. For each five-year period, we compute the cross-sectional average pricing errors and R^2s over all individual firms from the local, global, and pure local versions of the FFC six-factor model.

We first examine the differences between the local and global factor models. Panel A of Table 1.4 presents the differences in the time-series means of the cross-sectional average pricing errors and R^2s between the local and global factor models over all 271 five-year windows for each sample country. The differences of the pricing errors and R^2s are calculated as the values from the local factor model minus the values from the global factor model. In addition, the adjusted slope and the associated t-statistics from the regressions of those differences on time are reported, which demonstrates the time trends in the differences between the performances of the local and global factor models. It is known that when we estimate regressions using time-series variables, it is common for the residuals to have a time-series structure. Moreover, when we estimate the pricing errors and R^2s

[27] We also compare the performance of the equal-weighted factor models, where the weights $\omega_{j,t-1}$ are set to 1/9. Qualitatively similar results are obtained.

[28] We conduct a number of additional robustness checks through applying the alternative global factors to asset pricing tests, and obtain similar results.

for each individual stock over the 60-month rolling windows, the residuals from the individual stock regressions can be highly correlated. This violates the usual assumption of independent errors made in ordinary least squares regressions, which would result in unreliable t-statistics from the standard test. In order to adjust for the serial correlations in the residuals, we employ the Cochrane-Orcutt method to estimate the slopes and t-statistics.

[Table 1.4 inserted here]

As shown in Panel A of Table 1.4, the difference in the average pricing errors is uniformaly negative and the difference in the R^2s is uniformly positive for all sample countries, which indicates that the local factor model yields lower pricing errors and higher explanatory powers than the global factor model on average, which is consistent with the findings of [4]. Moreover, as shown by the adjusted slopes and t-statistics, the difference in the average pricing errors between the local and global factor models significantly decreased in France, Germany, Japan, and the U.S., and the difference in the average R^2s between the local and the global factor models significantly decreased in France, Germany, Hong Kong, Japan, and Switzerland.[29] The reduced differences in the performances of the local and global factor models for most of the sample countries imply that the local factors jointly became more globalized over time. Besides the convergence in the performance of the local and global factor models for the majority of the sample countries, we also find that the difference in the average pricing errors between the local and the global factor models does not exhibit a significant trend in the U.K., exhibit a increasing trend in Australia, Canada, Hong Kong, and Switzerland, while the difference in the average R^2s between the local and global factor models does not exhibit a significant trend in Australia, and significant upward trends in Canada, the U.K., and the U.S.

Why do we observe the convergence, divergence, or no trend in the differences in the

[29]When the difference in the average pricing errors is negative, a significant positive slope indicates convergence in the pricing errors from the local and global factor models. When the difference in the R^2s is positive, a significant negative slope indicates convergence in the explanatory power of the local and global factor models.

average pricing errors or R^2s across countries? There are many possible reasons. First, the time-trend could change over time. For example, we repeated the linear time trend analysis for Germany over the early sub-sample period and found that the difference in the average R^2s between the local and global factor models decreased significantly, which differs from the result obtained over the whole study period. Next, the performance of an asset pricing model can be substantially different across stocks within a country and may change over time (e.g., [27]). Moreover, the speed of the performance change for an asset pricing model could vary across stocks and countries. Therefore, the choice of an appropriate asset pricing model to explain stock returns could be dynamic within a country, across countries, and over time, depending on the market structure and degree of firm integration. Thus, taking average over all individual firms for the pricing errors and R^2s from the local and global factor models could result in convergent, divergent, or insignificant trends over time. Therefore, cross-sectional analyses of the performances of alternative factor models may provide further insights into the time trend results from a different prospective.[30]

In order to further compare the performances of the local, global, and pure local factor models in explaining individual stock returns and to examine the effect of globalization on the local factors, the time-series means of the cross-sectional average pricing errors and R^2s over all five-year periods from the alternative factor models are presented in Panel B of Table 1.4.

The pricing errors evaluate the regressions as asset pricing models. The absolute alpha for the local factor model is 1.73% on average, ranging from 1.10% for Switzerland to 2.50% for Canada. For the global factor model, the absolute alpha is 1.90% on average, ranging from 1.30% for Switzerland to 2.65% for Canada. For the pure local factor model, the absolute alpha is 1.99% on average, ranging from 1.40% for Switzerland to 2.74% for Canada. On average, the local factor model has the smallest pricing errors, followed by the global factor model and the pure local factor model. The local factor model delivers

[30]More detailed discussion is provided in subsection of Section 1.4 below.

more accurate average pricing than the global factor model for each country. The pure local factor model yields the highest average pricing errors in all sample countries, except Hong Kong, which indicates that the relatively more accurate pricing of the local factor model may be attributable to the globalization of the local factors.

The R^2s evaluate the factor models in terms of their effectiveness in explaining the time-series variations in the stock returns. The average R^2 from the local factor model is 0.25, ranging from 0.18 for the U.S. to 0.31 for Switzerland. For the global factor model, the average R^2 is 0.16, ranging from 0.13 for Hong Kong and Japan to 0.20 for Switzerland. For the pure local factor model, the average R^2 is 0.08, ranging from 0.02 for the U.S. to 0.15 for Japan. On average, the local factor model yields the highest R^2s, followed by the global factor model and the pure local factor model. The ranking is consistent in each country, except Japan. For Japan, the local factor model has the highest average R^2 (0.30), but the pure local factor model has an average R^2 of 0.15, which is higher than the average R^2 (0.13) from the global factor model. This implies that the Japanese local factors may not be highly globalized and the pure local factors may have an important function in asset pricing. For most sample countries, the substantially smaller R^2s from the pure local factor model compared with the R^2s from the local or global factor model indicate that the high explanatory power of the local factor model most likely arises from the effect of globalization.

In sum, the local factor model outperforms the global factor model in terms of pricing errors and explanatory powers in most cases, which is consistent with the findings in [4]. However, the pure local factor model yields the highest pricing errors and also the smallest explanatory powers in most cases. The poor performance of the pure local factor model provides a clue to the seemingly inconsistent results reported in the literature, that is, on the one hand, the local (country-specific) factors explain stock returns better than the global factors ([4]), but on the other hand, previous studies, such as [28] and [5], advocate models with both local and foreign factors, which is cognizant of the more integrated nature of world financial markets. The stronger performance of the local factor model compared with the global factor model is likely to be due, in part, to the globalization of the local factors.

1.4.2 Size-BE/ME and Size-MOM Sorted Portfolios

In this subsection, the dependent variables are the portfolio returns in excess of the U.S. risk-free rate. Due to the small number of firms for Australia, France, Germany, Hong Kong, and Switzerland in the early periods, 25 size-BE/ME (book-to-market) sorted and size-MOM (momentum) sorted portfolios are formed for four countries only, i.e., Canada, Japan, the U.K., and the U.S., over the whole sample period from July 1990 to December 2017. Then, nine size-BE/ME sorted and size-MOM sorted portfolios are formed for all nine sample countries over the shorter sample period from January 1996 to December 2017. Twenty-five portfolios are formed from the five-by-five independent sortings on size and BE/ME or MOM. Nine portfolios are formed from the three-by-three independent sortings on size and BE/ME or MOM. Then, the excess returns of each portfolio are regressed on the local, global, and pure local factors separately over the relevant sample periods.

Panel A of Table 1.5 presents the cross-sectional averages of the pricing errors and R^2s from the local, global, and pure local versions of the FFC six-factor model for 25 size-BE/ME sorted and size-MOM sorted portfolios. For the size-BE/ME sorted portfolios, the average pricing errors from the local factor model are 0.39%, 0.29%, 0.15%, and 0.10% for Canada, Japan, the U.K., and the U.S., respectively; 0.31%, 0.59%, 0.23%, and 0.38% from the global factor model; and 0.81%, 1.03%, 0.64%, and 0.58% from the pure local factor model. The local factor model yields smaller pricing errors than the global factor model for Japan, the U.K., and the U.S., but not for Canada, with the pure local factor model yielding the highest pricing errors in each of the four countries. For the size-BE/ME sorted portfolios, the average R^2s from the local factor model are 0.63, 0.75, 0.78, and 0.89 for Canada, Japan, the U.K., and the U.S., respectively; 0.48, 0.38, 0.50, and 0.71 from the global factor model; and 0.26, 0.40, 0.27, and 0.16 from the pure local factor model. The local factor model has higher average R^2s than the global factor model, but the pure local factor model yields much smaller average R^2s than the global factor model. We find very similar performance results using the 25 size-MOM sorted portfolios; furthermore, the relative performance of the three versions of the FFC six-factor model in explaining the portfolio returns is essentially the same as that in explaining the individual stock returns. The poor performance of the pure local factor model in all sample countries reflects the globalization of the local factors.

[Table 1.5 inserted here]

In order to evaluate the hypothesis that the model intercepts from the 25 size-BE/ME sorted or size-MOM sorted portfolios are jointly equal to zero, we conduct GRS tests. The GRS test statistics (F-stat) and p-values (p-val) are provided in Panel B of Table 1.5. For the size-BE/ME sorted portfolio regressions, the F-statistics from three versions of the FFC six-factor model fail to reject the null hypothesis that the intercepts are jointly equal to zero for all specifications at the 10% significance level in Canada and the U.K., but reject the null hypothesis for all specifications in the U.S. at the 5% significance level. For Japan, the F-statistics from the local and global versions of the FFC six-factor model fail to reject the null hypothesis at the 10% significance level, but the F-statistics from the pure local version of the model rejects the null hypothesis at the 1% level of significance. In contract, for the size-MOM sorted portfolio regressions, the F-statistics from three versions of the FFC six-factor model strongly reject the null hypothesis that the intercepts are jointly equal to zero for all specifications at the 5% significance level for all countries. The F-statistics can also be used to evaluate the regressions as asset pricing models. The smaller the F-statistics, the better the asset pricing model is. In terms of the GRS F-test statistics, a consistent ranking does not exist between the local and the global factor models, but the pure local factor model consistently performs most poorly among the three factor models.

In order to expand the comparison of the local, global, and pure local versions of the FFC four-factor model beyond the four countries we investigated, we repeat the preceding analysis for the nine size-BE/ME and size-MOM sorted portfolios from all nine sample countries over the period from January 1996 to December 2017. The cross-sectional average pricing errors and R^2s, and the GRS F-test results, are provided in Table 1.6. As can be seen from the table, the results of the pricing errors and explanatory powers of the three versions of the FFC six-factor model for nine sample countries are qualitatively similar to those reported in Table 1.5 for the four sample countries, but the results of GRS tests demonstrate some differences across countries. For example, for the nine size-BE/ME sorted portfolio regressions, the F-statistics from three versions of the FFC six-factor model reject the null hypothesis that the intercepts are jointly equal to zero for all specifications at the 10% significance level for Australia, France, Hong Kong, and Japan. For the other countries, the F-statistics from some versions of the six-factor model fails to reject the null hypothesis at the 10% level of significance, such as the global version of the model in Canada, Germany, Switzerland, and the U.K. For the size-MOM sorted portfolio regressions, the F-statistics from three versions of the FFC six-factor model reject the null hypothesis that the intercepts are jointly equal to zero for all specifications at the 1% significance level for all specifications in all countries except only the local version of the model in the U.S.

[Table 1.6 inserted here]

Overall, the large GRS F-statistics in most cases indicate that all three versions of the FFC six-factor model are rejected as asset pricing models. However, the local factor model has smaller pricing errors and larger explanatory powers than the global factor model in most instances. The pure local factor yields the poorest performance in terms of the cross-sectional average pricing errors and R^2s. These findings imply that the superior performance of the local factor model vis-a-vis the global factor model in terms of a more accurate pricing and greater explanatory power is partly due to the globalization of the local factors.

1.4.3 Size-R^2 Sorted Portfolios

As shown above, the local factor model is more useful for explaining the time-series variations in individual stock and portfolio returns than the global factor model, due to the globalization of the local factors. However, none of the three factor models can fully explain the average returns. Can this relatively poor performance of the asset pricing models be related to the possibly heterogeneous integration? That is, could the global factor model be useful for pricing highly globally integrated stocks, while the local factor model or pure local factor model be more useful for pricing more segmented stocks?

In order to examine the ability of the local, global, and pure local factor models to explain the stock returns with heterogeneous degrees of integration, we form 25 portfolios from five independent sortings on firm size and firm-level R^2s estimated from the global factor model for Canada, Japan, the U.K., and the U.S. over the period from July 1995 to December 2017.[31] Each month, beginning from July 1995, we use the previous 60-month observations of the individual stock returns in the global factor model regression for each country to estimate the R^2s, and then form 25 size-R^2 sorted portfolio monthly returns from five independent sorts on firm size of the previous month and five independent sorts on R^2s from the global factor model regressions. Portfolio 1 represents the small and segmented portfolio, and portfolio 25 represents the large and integrated portfolio. A stock is included in the construction of the portfolios only if the stock has all 60 monthly observations available during the five-year rolling window. The pricing errors and R^2s for each portfolio from the local, global, and pure local factor model regressions for Canada, Japan, the U.K., and the U.S. are provided in Table 1.7 and are plotted in Figure 1.2.

[Table 1.7 and Figure 1.2 inserted here]

As presented in Table 1.7 and Figure 1.2, the pure local factor model yields the highest pricing errors in most cases. The local factor model has the highest R^2s, followed by the

[31] We also repeat the analysis for all nine sample countries, with 9 portfolios formed from three-by-three independent sortings on firm size and R^2 from January 2001 to December 2017. Similar results are obtained.

global factor model and the pure local factor model for most portfolios in Canada, the U.K., and the U.S. In Japan, the local and pure local factor models have much higher R^2s than the global factor model in most cases, which indicates that Japanese local factors have important function in explaining the portfolio returns. However, it is noteworthy that for some large and integrated portfolios in Japan, the R^2s (pricing errors) from the global factor model tend to be higher (smaller) than those from the pure local factor model. These results imply that although the pure local factors might have an important function in asset pricing in Japan, the global factor model can outperform the pure local factor model for the most integrated portfolios.

In addition, the pricing errors of the global factor model exhibit a downward tendency from portfolio 1 to portfolio 25 for Canada and the U.S., no discernable trend for Japan and the U.K., and the R^2s of the global factor model exhibit a upward trend from portfolio 1 to portfolio 25 for all countries. Moreover, the differences in R^2s between the local and the global factor models exhibit a downward tendency from portfolio 1 to portfolio 25 for all countries. For example, the differences in the R^2s are 0.19, 0.64, 0.28, and -0.08 for the small and segmented portfolio (portfolio 1) in Canada, Japan, the U.K., and the U.S., respectively, and 0.13, 0.34, 0.14, and 0.06 for portfolio 25. Pricing errors are difficult to compare. In Japan, the global factor model yields the smallest pricing errors for the small or segmented portfolios, but it has a small explanatory power. The downward tendency in the differences of the explanatory power between the local and the global factor models from portfolio 1 to portfolio 25 indicates that the global factors have a more important function in asset pricing for the more integrated portfolios.[32]

As depicted in Figure 1.2, for the large and segmented portfolio (portfolio 21), there is a sharp decrease in the R^2s from the local and global factor model regressions. This might result from the small number of firms within the portfolio, compared with other portfolios.

[32] In the U.S., the R^2s of the small or segmented portfolios (portfolios 1 and 2) from the global factor model are higher than the R^2s from the local factor model, and the pricing errors from the global factor model are smaller than those from the local factor model.

The small number of firms within portfolio 21 also indicates that there are relatively few firms that are large and also segmented, which is consistent with the previous literature that large firms, which have high liquidity, lower cost of information, few barriers and good governance, tend to be more integrated (e.g., [29], [30]).

It is also noted that the pricing errors for all factor models are particularly large in portfolio 1 through portfolio 3, which are predominantly comprised of segmented stocks. Compared with the other portfolios, there are a sufficient number of firms within these three portfolios, which implies that there is enough number of firms for the diversification of idiosyncratic risk. Therefore, the large pricing errors in these portfolios indicate a possible omitted risk factor beyond the FFC six factors.

Overall, the results from the size-R^2 sorted portfolios confirm the preceding findings that the stronger performance of the local factor model arises from the globalization of the local factors, and the pure local factor model tends to perform poorly, except for Japan. Our results in this section also indicate that the global factors are more important in asset pricing for the more integrated portfolios, and the performance of an asset pricing model relates to the degree of firm integration.

1.4.4 Variance Decomposition of Portfolio Returns

The findings from the previous subsections imply that, in partially integrated stock markets, the global, local, and idiosyncratic components may all matter in pricing within each country, which reflects the firm attributes and the degree of globalization at the firm-level. Therefore, how large is the pricing effect from each component? In order to simultaneously compare the effects of the global factors, pure local factors, and idiosyncratic components on the variance of the portfolio returns, we decompose the variance of each size-R^2 sorted portfolio return into three components that are explained separately by the global factors, pure local factors, and idiosyncratic components.

For each country, we first regress the excess return on each size-R^2 sorted portfolio on the global and pure local factors. That is, we have

$$\begin{aligned}R_{p,i,t} &= \alpha_{p,i} + \beta_{p,i}^{GM} GMKT_t + \beta_{p,i}^{GS} GSMB_t + \beta_{p,i}^{GH} GHML_t \\ &+ \beta_{p,i}^{GO} GMOM_t + \beta_{p,i}^{GR} GRMW_t + \beta^{GC} GCMA_t \\ &+ \beta_{p,i}^{LM*} LMKT_{i,t}^* + \beta_{p,i}^{LS*} LSMB_{i,t}^* + \beta_{p,i}^{LH*} LHML_{i,t}^* + \beta_{p,i}^{LO*} LMOM_{i,t}^* \\ &+ \beta^{LR*} LRMW_{i,t}^* + \beta^{LC*} LCMA_{i,t}^* + \epsilon_{p,i,t},\end{aligned} \tag{4}$$

where $R_{p,i,t}$ is the excess return of the p-th size-R^2 sorted portfolio from the i-th country in month t; $LMKT_{i,t}^*, LSMB_{i,t}^*, LHML_{i,t}^*$, $LMOM_{i,t}^*$, $LRMW_{i,t}^*$, and $LCMA_{i,t}^*$ are the pure local MKT, SMB, HML, MOM, RMW, and CMA factors of the i-th country in month t, respectively; and "*" in the superscripts of β and the local factors indicates the pure local coefficients or factors.

By construction, the pure local factors are orthogonal to the global factors. Then we compute the proportions of the portfolio return variance explained by the global factors, the pure local factors, and the residual component in the regressions in equation (4). The proportion from the residual component in the regressions is the proportion accounted for by the idiosyncratic components.

$$Q_{p,i}^G + Q_{p,i}^{L*} + Q_{p,i}^I = 1, \tag{5}$$

where $Q_{p,i}^G$, $Q_{p,i}^{L*}$, and $Q_{p,i}^I$ are the proportions of the p-th portfolio return variance accounted for by the global factors, pure local factors, and idiosyncratic component in the i-th country, respectively,

$$Q_{p,i}^{G} = \frac{Var(\beta_{p,i}^{GM}GMKT + \beta_{p,i}^{GS}GSMB + \beta_{p,i}^{GH}GHML + \beta_{p,i}^{GO}GMOM + \beta_{p,i}^{GR}GRMW + +\beta_{p,i}^{GC}GCMA)}{Var(R_{p,i})}, \tag{6a}$$

$$Q_{p,i}^{L*} = \frac{Var(\beta_{p,i}^{LM*}LMKT_i^* + \beta_{p,i}^{LS*}LSMB_i^* + \beta_{p,i}^{GH*}LHML_i^* + \beta_{p,i}^{GO*}LMOM_i^* + \beta_{p,i}^{GR*}LRMW_i^* + \beta_{p,i}^{GC*}LCMA_i^*)}{Var(R_{p,i})}, \tag{6b}$$

$$Q_{p,i}^{I} = 1 - Q_{p,i}^{G} - Q_{p,i}^{L*}. \tag{6c}$$

Figure 1.3 plots the proportions explained by the global, pure local, and idiosyncratic components for the 25 size-R^2 sorted portfolios in Canada, Japan, the U.K., and the U.S.[33]

[Figure 1.3 inserted here]

As depicted in Figure 1.3, for most portfolios, the global factors have the largest explanatory power, followed by the idiosyncratic components and the pure local factors in Canada, the U.K., and the U.S, which is consistent with the preceding results. Specifically, the global factors account for 44%, 47%, and 74% of the variance, on average, for 25 size-R^2 sorted portfolios in Canada, the U.K., and the U.S., respectively, followed by the idiosyncratic components accounting for 37%, 29%, and 18%, respectively, and the pure local factors accounting for 18%, 24%, and 8%, respectively. In Japan, the pure local factors account for a much greater proportion of the portfolio return variance (59%) than the global factors (27%) and the idiosyncratic components (13%), which confirms the preceding finding that Japanese country-specific factors are important in asset pricing. Figure 1.3 also shows that the global factors have the largest explanatory power for portfolio 24 and 25 in Japan, which is similar to the preceding finding that indicates that the returns of the highly globalized stocks are primarily driven by global factors.

[33]The detailed proportions explained by the global, pure local, and idiosyncratic components are provided in Table A.3 in the Appendix.

In addition, the idiosyncratic components explain a large proportion of the small portfolio returns (portfolios 1 to 5) , and large and segmented portfolio returns (portfolios 21 and 22), particularly in Canada and the U.K., which partly explains the large pricing errors for portfolios 1 to 3 in Figure 3 and Table 1.7. This implies that the idiosyncratic risk may need to be adequately considered for pricing the small and segmented stocks/portfolios. Further studies are needed in order to examine the role of the idiosyncratic risk, whether and how the idiosyncratic component can help achieve more accurate pricing.

Overall, our results from the variance decomposition of the size-R^2 sorted portfolio returns are supportive of our earlier finding that the strong performance of the local factor model is attributable to the globalization of the local factors. Moreover, the global factors account for the portfolio returns' variance more than the pure local and idiosyncratic components for large and integrated stocks in all countries, while the idiosyncratic components account for large proportions of the small or large and segmented stock returns, and thus may not be ignored in pricing those stocks.

1.5 Griffin Revisited, 1981–1995

In this section, we replicate the [4] study using the same sample period and the same Fama and French (FF) three-factor model, i.e., the MKT, SMB, and HML factors, but considering the effect of the globalization of the local factors in interpreting the key findings. While our findings provided in the previous sections already offer clear insight into the international asset pricing "puzzle", it remains useful to replicate his study in order to provide further credence to this insight.

As [4] did, we evaluate the performance of the three alternative versions of the FF three-factor model, i.e., the domestic, world, and international factor models, in pricing individual stocks and BE/ME sorted portfolios for four major stock markets, i.e., Canada, Japan, the U.K., and the U.S. over the sample period from January 1981 to December 1995.[34] The domestic factor model is a country-specific (i.e., "local") version of the FF three-factor model comprising the local MKT, SMB, and HML factors in our study; the world factor model is the "global" factor model in our study, comprising the global MKT, SMB, and HML factors; and the international factor model adds foreign components to the domestic factor model, thus comprising three local factors plus

[34] Data for Canada, Japan, and the U.K. in this section are downloaded from DataStream. See Table A.1.

three international factors constructed from foreign markets in the sample. In addition to the three versions of the FF factor model mentioned above, we further consider the pure local factor model introduced previously to assist in interpreting the effect of globalization on the local factors in explaining the stock returns. Again, following [4], both weighted (value-weighted) and unweighted (equal-weighted) factor regressions are estimated in order to compare the alternative factor models.

In order to replicate [4], we estimate the individual stock regressions over five-year rolling windows, beginning in January 1981. The time-series means of the cross-sectional averages of the pricing errors and R^2s for all individual stocks over all rolling windows are presented in Panel A of Table 1.8. For the 25 size—BE/ME sorted portfolios, on the other hand, we estimate each portfolio regression over the entire sample period of January 1981 – December 1995. The cross-sectional average pricing errors and R^2s from each of the four alternative factor models are reported in Panel B of Table 1.8, with the associated GRS F-test results reported in Panel C of Table 1.8.

[Table 1.8 inserted here]

As can be seen from Panel A of Table 1.8, the local (domestic) factor model outperforms the global (world) factor model in terms of both the average pricing errors and R^2s, while the international factor model, which adds foreign factors to the local factor model, leads to greater pricing errors with very similar R^2s. These findings are fully consistent with those reported in Griffin (2002). In particular, the average pricing error from the weighted local (global) factor regressions is 1.46% (1.58%), 0.94% (1.30%), 1.16% (1.29%), and 1.23% (1.48%) for Canada, Japan, the U.K., and the U.S., respectively, and 1.53%, 1.02%, 1.30%, and 1.42%, respectively, from the weighted international factor model. The average R^2 from the weighted local (global) factor model is 0.18 (0.09), 0.43 (0.27), 0.31 (0.15), and 0.21 (0.14) for Canada, Japan, the U.K., and the U.S., respectively, and 0.19, 0.43, 0.31, and 0.21, respectively, from the weighted international factor model. The pure local factor model yields the highest pricing errors and lowest R^2s among the four factor models: the average pricing error (R^2) is 1.78% (0.09), 1.53% (0.14), 1.75% (0.15), and 1.51% (0.07) for Canada, Japan, the U.K., and the U.S., respectively. As can be seen from Panel B of Table 1.8, we obtain very similar results for the 25 size-BE/ME sorted portfolios. In addition, the GRS test results provided in Panel C of Table 1.8 indicate that none of the factor models can fully explain the portfolio returns, but the local factor model performs better than other models and the pure local factor model performs most poorly. It is also noted that the unweighted factor regressions produce largely the same results as those from the weighted factor regressions.

Overall, our findings in this section are qualitatively very similar to those of [4] who documented that the local (domestic) FF three-factor model outperforms the world and international versions of the FF three-factor models in explaining the average stock returns. However, we additionally show that the pure local version of the FF three-factor model performs most poorly, implying that the superior performance of the local factor model can be attributable to the globalization of the local factors. Despite the different sample period and factor model used in this section, most findings are fundamentally the same as those reported in the previous sections where the FFC four-factor model is employed over different periods.[35]

1.6 Conclusion

Motivated by the large study of globalization focused on the market factors, we extend the scope of the analysis to the other FFC local factors, which explains the intriguing finding of integrated international asset pricing ([4], [6]) that the local version of the Fama-French three-factor model outperforms the global version despite the evident globalization of financial markets in recent decades, we conjectured that this finding might be attributable to the globalization of the local factors. Using the R^2 method to measure the degree of global integration, we found that most of our sample Fama-French five local factors, as well as the Carhart local momentum factors, are significantly globalized. It is remarkable that almost all of our sample FFC local factors are significantly globalized; that is, few local factors are "purely local" over the whole sample period. Furthermore, our assessments of competing FFC factor models indicate that the local factor model outperforms the global factor model, which is broadly consistent with the finding of integrated international asset pricing, but this is due to the globalization of the local factors, as evidenced by our finding that the global factor model clearly outperforms the pure local factor model in most cases.

However, our observation that the relative performances of the competing factor models are

[35] We also repeat our analysis using the five-factor model proposed by [7], and we find that the results are similar to the results of the FFC four-factor model. Detailed results are available upon request.

not steady but rather vary across our sample portfolios sorted on different attributional dimensions, e.g., size–BE/ME and size–R^2, points to a realistic possibility, i.e., the particular factors relevant for pricing may vary across the firm size, degree of firm globalization, and other attributes, within a country. Furthermore, the variance decomposition results lead us to conjecture that the relative pricing power of the global, (pure) local, and idiosyncratic factors may vary across firms, within and across countries. This complicated situation increases the difficulty of selecting one particular version of FF or FFC factor model over alternative versions for the purpose of cost of capital estimations and portfolio performance evaluations. Furthermore, in partially integrated world financial markets, stocks may be priced simultaneously by both the global and local factors, with varying relative weights between them, depending on the firm attributes and the degree of globalization. This unsettled situation calls for further research.

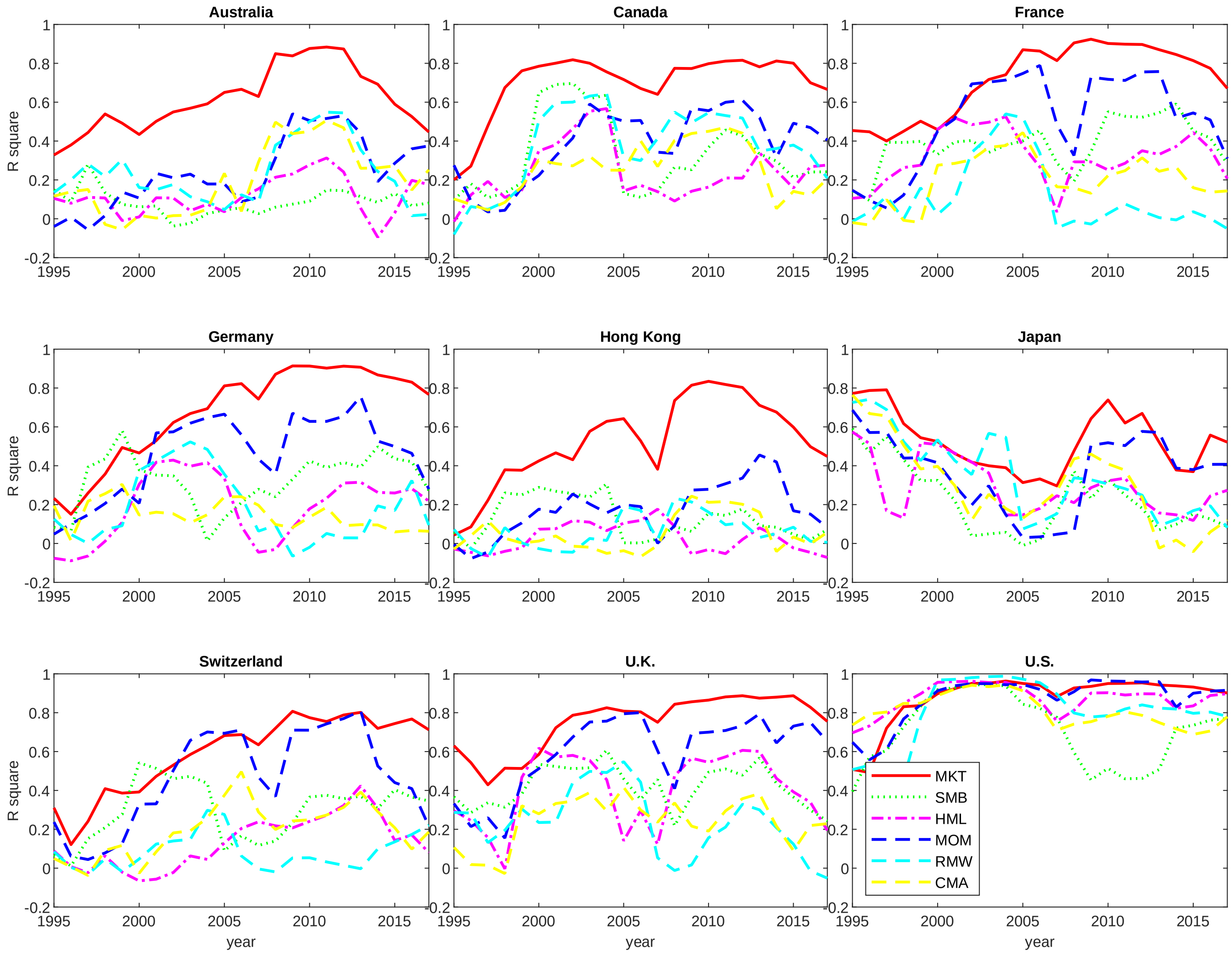

Figure 1.1: The time-trends in globalization for the local factors of the sample countries.

For each sample country, each of the six local factors (MKT, SMB, HML, MOM, RMW, CMA) is regressed on the global factors over five-year rolling windows during the whole sample period of July 1990 – December 2017. The global factors are constructed using the market value-weighted averages of the local factors of Australia, Canada, France, Germany, Hong Kong, Japan, Switzerland, the U.K., and the U.S. The time-series R^2s from the regressions are plotted for each sample country. The x-axis represents the time (ending month of each five-year window), and the y-axis represents the degree of globalization, measured by R^2.

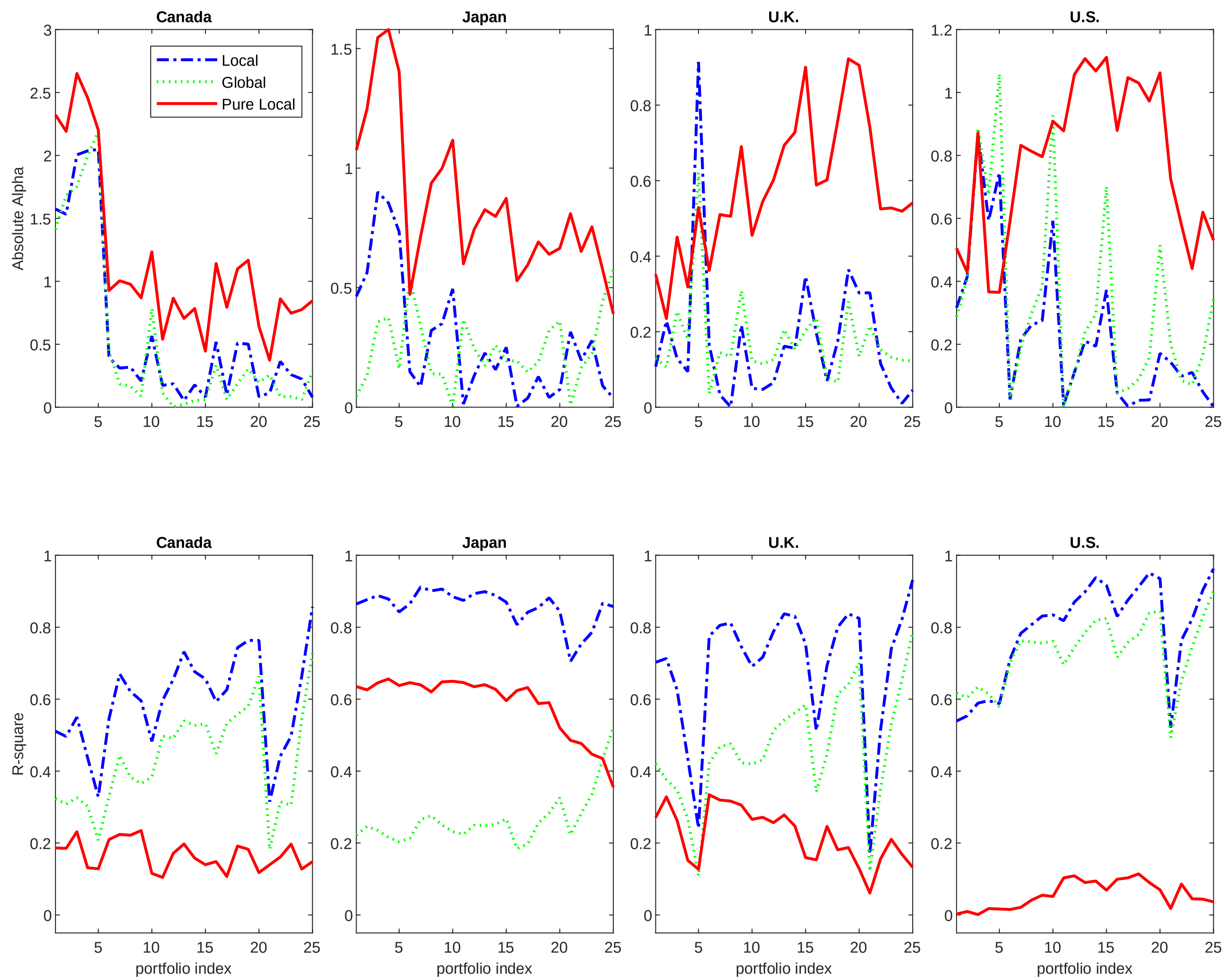

Figure 1.2: Pricing errors and R^2s for the 25 size-R^2 sorted portfolios.

For Canada, Japan, the U.K., and the U.S., the 25 size-R^2 sorted portfolios' excess returns are regressed on the local, global, and pure local factors over the period of July 1995 – December 2017. The upper panel plots the pricing errors and the lower panel plots the R^2s from the regressions on the local, global, and pure local factors. The x-axis represents the portfolio index, from the small and segmented portfolio (portfolio 1) to the large and integrated portfolio (portfolio 25). The y-axis represents the pricing error in the upper panel and the explanatory power (R^2) in the lower panel.

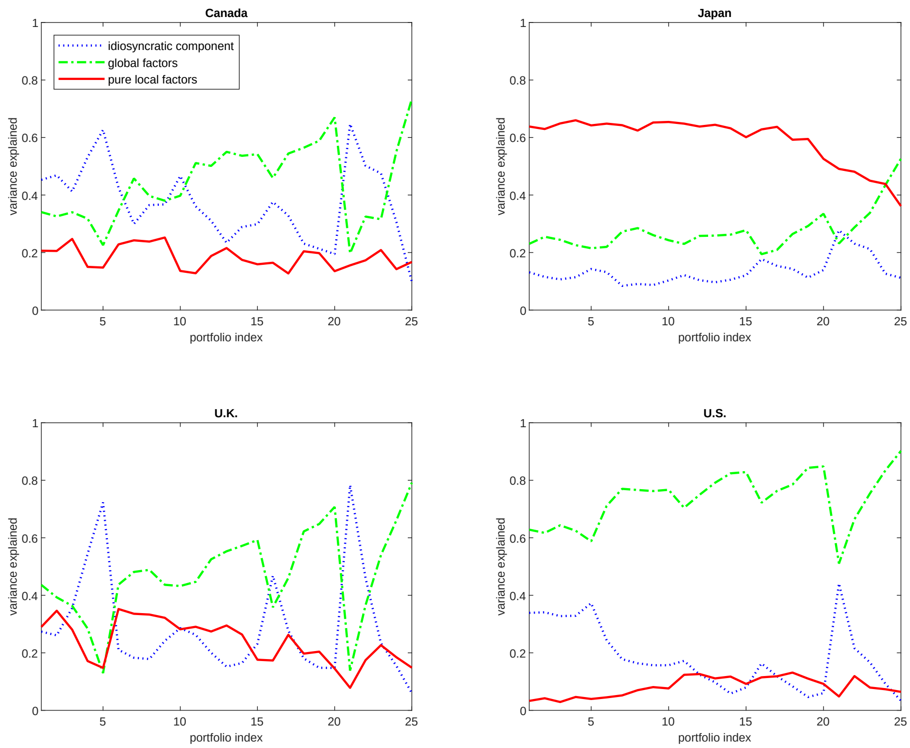

Figure 1.3: Variance decomposition of the size-R^2 sorted portfolios.
For Canada, Japan, the U.K, and the U.S., the 25 size-R^2 sorted portfolios' excess returns are regressed on the global and pure local factors over the period of July 1995 – December 2017. The proportions of each portfolio return variance explained by the global factors, pure local factors, and idiosyncratic components are plotted for each sample country. The solid red line plots the proportion of the variance explained by the pure local factors, the dashed green line plots the proportion of the variance explained by the global factors, and the blue dotted line plots the proportion of the variance explained by the idiosyncratic components. The x-axis represents the portfolio index, from the small and segmented portfolio (portfolio 1) to the large and integrated portfolio (portfolio 25). The y-axis represents the explained proportion.

Table 1.1: Summary statistics.

For each sample country, (i) the average number and size of the sample firms, (ii) the mean, median, standard deviation (Std), maximum (Max), and minimum (Min) of each of the six local factors (MKT, SMB, HML, MOM, RMW, CMA), and (iii) the Sharpe ratio (SHP) of each of the six local factors are reported, with the corresponding summary statistics for the value-weighted global factors. The value-weighted global factors are constructed using the market value-weighted averages of the local factors of Australia, Canada, France, Germany, Hong Kong, Japan, Switzerland, the U.K., and the U.S. The returns of all local factors and global factors are in U.S. dollars and are expressed in percent per month for the period of July 1990 – December 2017. The market (MKT) factor of each country is the market return of the country in excess of the one-month U.S. Treasury bill rate.

Country	Average number of sample firms	MKT Mean	MKT Median	MKT Std	MKT Max	MKT Min	MKT SHP	SMB Mean	SMB Median	SMB Std	SMB Max	SMB Min	SMB SHP	HML Mean	HML Median	HML Std	HML Max	HML Min	HML SHP
Australia	513	0.78	1.10	5.81	17.41	-26.78	0.13	0.46	0.40	3.40	11.87	-10.09	0.13	0.53	0.47	2.92	12.27	-9.82	0.18
Canada	692	0.68	0.93	5.19	20.77	-27.09	0.13	0.36	0.32	4.17	22.51	-14.06	0.09	0.27	0.50	4.58	18.38	-20.65	0.06
France	460	0.51	0.74	5.55	15.77	-23.15	0.09	0.04	-0.02	3.30	10.02	-12.62	0.01	0.29	0.18	3.10	10.44	-20.45	0.09
Germany	466	0.47	0.68	5.60	18.38	-20.57	0.08	-0.14	-0.41	3.49	8.77	-13.78	-0.04	0.37	0.23	2.88	13.19	-9.28	0.13
Hong Kong	571	0.87	0.96	7.04	28.31	-30.67	0.12	0.08	-0.21	6.12	27.12	-27.50	0.01	0.42	0.18	3.87	20.69	-12.92	0.11
Japan	2693	0.11	0.24	5.77	25.52	-17.58	0.02	0.02	-0.05	3.37	13.98	-11.93	0.01	0.53	0.64	2.74	11.42	-11.78	0.19
Switzerland	163	0.70	0.75	4.69	15.65	-17.27	0.15	-0.03	0.05	3.50	19.85	-12.69	-0.01	0.04	-0.02	2.79	8.74	-13.74	0.01
U.K.	985	0.47	0.56	4.55	14.84	-20.68	0.10	0.24	0.28	3.71	15.32	-13.11	0.06	0.18	0.12	3.42	15.69	-14.36	0.05
U.S.	4590	0.69	1.14	4.20	11.36	-16.96	0.17	0.16	0.12	3.20	21.45	-16.44	0.05	0.23	-0.01	3.16	13.41	-13.54	0.07
Value-weighted global factors:		0.49	0.49	4.22	11.08	-18.50	0.11	0.12	0.06	2.27	14.44	-11.05	0.05	0.30	0.11	2.30	9.95	-11.45	0.13

Country	Average size of sample firms ($M)	MOM Mean	MOM Median	MOM Std	MOM Max	MOM Min	MOM SHP	RMW Mean	RMW Median	RMW Std	RMW Max	RMW Min	RMW SHP	CMA Mean	CMA Median	CMA Std	CMA Max	CMA Min	CMA SHP
Australia	1063.87	1.67	1.78	4.29	14.33	-18.88	0.39	0.07	0.11	3.64	13.93	-11.36	0.02	0.34	0.29	2.90	8.94	-11.22	0.12
Canada	1152.32	1.53	1.79	6.01	18.46	-22.99	0.26	0.64	0.78	5.48	19.36	-15.97	0.12	0.56	0.31	3.79	13.97	-9.08	0.15
France	2789.29	0.88	1.15	4.31	20.67	-22.75	0.21	0.13	0.13	2.14	6.98	-8.64	0.06	0.36	0.27	2.53	9.98	-9.79	0.14
Germany	3062.14	1.10	1.16	5.05	21.30	-25.27	0.22	0.08	0.19	2.61	8.29	-11.17	0.03	0.09	-0.01	2.98	13.48	-10.68	0.03
Hong Kong	1155.41	0.74	1.18	5.93	15.45	-34.02	0.12	0.62	0.81	4.81	18.20	-17.11	0.13	-0.20	-0.17	3.45	12.31	-15.96	-0.06
Japan	1367.02	0.11	0.65	4.55	15.12	-30.38	0.02	0.04	0.21	2.09	9.12	-7.78	0.02	0.06	-0.06	2.14	8.72	-7.48	0.03
Switzerland	3265.62	0.74	0.67	4.77	23.89	-27.95	0.16	0.25	0.48	2.78	7.96	-9.74	0.09	0.17	0.02	3.39	15.66	-9.33	0.05
U.K.	2097.93	1.27	1.38	4.66	16.02	-35.80	0.27	0.25	0.23	2.26	7.04	-11.48	0.11	0.17	0.27	2.38	10.98	-13.33	0.07
U.S.	2868.04	0.44	0.53	4.78	18.08	-34.82	0.09	0.37	0.38	2.76	13.27	-18.81	0.13	0.22	0.03	1.94	9.32	-6.97	0.11
Value-weighted global factors:		0.57	0.84	3.87	13.96	-28.05	0.15	0.28	0.33	1.89	8.61	-12.60	0.15	0.18	0.03	1.38	7.02	-4.13	0.13

Table 1.2: Globalization of local factors, measured by R^2.

For each sample country, each of the six local factors (MKT, SMB, HML, MOM, RMW, CMA) is regressed on the global factors over five-year rolling windows for the whole sample period of July 1990 – December 2017, and the two evenly divided sub-sample periods of July 1990 – March 2004 and April 2004 – December 2017. The global factors are constructed from the market value-weighted averages of the local factors of Australia, Canada, France, Germany, Hong Kong, Japan, Switzerland, the U.K., and the U.S. The time-series average R^2 from the regressions for each local factor is reported. The difference between the average R^2s from the two sub-sample periods for each local factor is also reported. In addition, we formally test the significance of the average R^2 and whether the average R^2 from the recent sub-sample period is larger than the average R^2 from the earlier sub-sample period for each local factor. We then formally test whether R^2s change over time. The adjusted t-statistics are reported, which considers the serial correlation of R^2s. "*", "**", and "***" indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Whole sample period 1990-2017							
Country	Local Factor	MKT	SMB	HML	MOM	RMW	CMA
Australia	Mean	0.616	0.091	0.113	0.235	0.231	0.197
	t-stat	9.486***	4.188***	3.933***	3.503***	4.311***	2.720***
	Slope	0.001	0.000	0.000	0.002	0.000	0.001
	t-stat	2.404**	−0.080	1.107	4.367***	0.815	2.562**
Canada	Mean	0.694	0.329	0.227	0.392	0.375	0.251
	t-stat	12.040***	5.258***	4.938***	5.861***	5.437***	5.129***
	Slope	0.001	0.000	0.000	0.002	0.001	0.001
	t-stat	1.853*	0.082	0.436	2.925***	1.213	1.065
France	Mean	0.713	0.391	0.320	0.512	0.123	0.193
	t-stat	9.128***	10.309***	7.552***	5.864***	1.993**	4.262***
	Slope	0.002	0.001	0.000	0.002	0.000	0.001
	t-stat	3.840***	2.735***	0.497	1.980**	−0.955	1.210
Germany	Mean	0.683	0.324	0.189	0.463	0.191	0.161
	t-stat	7.122***	7.914***	3.264***	6.191***	2.882***	8.148***
	Slope	0.003	0.001	0.001	0.002	0.000	0.000
	t-stat	4.825***	1.325	1.547	2.072**	−0.312	−2.641***
Hong Kong	Mean	0.531	0.125	0.030	0.176	0.053	0.058
	t-stat	6.274***	3.649***	1.331	3.782***	1.920*	1.850*
	Slope	0.002	0.000	0.000	0.001	0.000	0.000
	t-stat	2.961***	−1.175	−0.397	2.776***	2.134**	1.288
Japan	Mean	0.527	0.229	0.290	0.367	0.355	0.311
	t-stat	10.176***	3.841***	7.732***	5.529***	4.790***	4.060***
	Slope	−0.001	−0.001	−0.001	0.000	−0.002	−0.002
	t-stat	−1.394	−2.857***	−4.146***	−0.374	−8.302***	−4.351***
Switzerland	Mean	0.589	0.290	0.129	0.471	0.082	0.193
	t-stat	7.457***	6.089***	2.431**	5.190***	2.921***	3.735***
	Slope	0.002	0.001	0.001	0.002	0.000	0.001
	t-stat	6.291***	1.087	3.698***	2.055**	0.914	2.647***
U.K.	Mean	0.751	0.421	0.397	0.592	0.244	0.239
	t-stat	13.222***	15.379***	7.510***	7.759***	4.806***	5.673***
	Slope	0.001	0.000	0.001	0.002	−0.001	0.001
	t-stat	4.838***	0.148	1.211	3.290***	−1.820*	1.211
U.S.	Mean	0.879	0.702	0.871	0.876	0.807	0.810
	t-stat	19.576***	10.579***	37.477***	20.890***	13.840***	26.630***
	Slope	0.001	0.000	0.000	0.001	0.001	0.000
	t-stat	2.182**	−0.604	0.646	2.071**	1.101	−1.822*

Panel B: Two sub-sample periods

Country	Local Factor	July 1990-March 2004 MKT	SMB	HML	MOM	RMW	CMA	April 2004- December 2017 MKT	SMB	HML	MOM	RMW	CMA	Sub-sample comparisons Δ MKT	Δ SMB	Δ HML	Δ MOM	Δ RMW	Δ CMA
Australia	Mean	0.466	0.098	0.076	0.096	0.189	0.046	0.744	0.115	0.157	0.420	0.334	0.354	0.279	0.018	0.080	0.324	0.145	0.308
	t-stat	12.722***	2.631***	9.525***	1.756*	14.198***	1.508	10.656***	14.214***	2.962***	7.794***	3.593***	5.486***	2.706***	0.453	1.737**	2.999***	1.754**	7.325***
	Slope	0.002	−0.002	0.000	0.003	0.000	−0.002	−0.004	0.000	−0.002	−0.003	−0.005	−0.004						
	t-stat	7.486***	−5.524***	−1.434	12.349***	−1.093	−3.938***	−5.662***	0.903	−2.192**	−4.864***	−4.105***	−8.418***						
Canada	Mean	0.601	0.374	0.243	0.230	0.289	0.175	0.776	0.324	0.214	0.506	0.428	0.304	0.175	−0.050	−0.029	0.276	0.139	0.129
	t-stat	4.940***	2.710***	2.643***	3.348***	2.068**	2.915***	47.123***	11.459***	9.396***	13.227***	8.848***	3.873***	1.349*	−0.307	−0.398	2.665***	0.739	0.935
	Slope	0.007	0.008	0.006	0.004	0.008	0.003	−0.001	−0.001	0.001	−0.002	−0.003	−0.004						
	t-stat	5.858***	8.285***	21.922***	3.266***	15.554***	8.374***	−1.913*	−1.924*	4.267***	−6.165***	−8.058***	−5.572***						
France	Mean	0.505	0.322	0.332	0.329	0.122	0.130	0.855	0.475	0.326	0.625	0.017	0.209	0.350	0.154	−0.006	0.296	−0.105	0.079
	t-stat	11.094***	5.968***	4.106***	2.731***	2.276**	1.654	26.123***	17.003***	14.347***	10.636***	1.270	10.109***	4.487***	2.710***	−0.089	1.663**	−1.587	0.847
	Slope	0.003	0.003	0.005	0.007	0.004	0.005	−0.002	0.000	0.001	−0.003	−0.001	0.000						
	t-stat	3.838***	4.408***	11.012***	7.238***	4.513***	11.796***	−6.097***	−0.368	1.539	−4.308***	−1.584	−0.793						
Germany	Mean	0.429	0.329	0.154	0.310	0.222	0.181	0.879	0.398	0.231	0.576	0.100	0.115	0.450	0.069	0.078	0.266	−0.123	−0.066
	t-stat	5.072***	5.409***	1.390	2.755***	2.253**	10.390***	47.149***	30.685***	6.282***	11.444***	1.666*	5.911***	4.408***	1.381*	0.877	1.678**	−3.054	−2.949
	Slope	0.005	0.002	0.006	0.007	0.006	0.000	−0.001	0.000	0.002	−0.003	0.003	−0.001						
	t-stat	11.157***	1.349	11.577***	19.363***	6.522***	−0.210	−4.931***	0.602	2.342**	−3.292***	9.347***	−10.117***						
Hong Kong	Mean	0.324	0.194	0.031	0.088	−0.014	0.028	0.707	0.100	−0.014	0.280	0.096	0.129	0.383	−0.093	−0.045	0.192	0.110	0.101
	t-stat	3.775***	3.640***	0.867	1.516	−3.095***	2.084**	11.355***	4.350***	−1.033	7.387***	3.003***	2.544**	2.639***	−1.299	−1.236	2.348***	3.581***	2.217**
	Slope	0.005	0.003	0.002	0.003	0.000	0.000	−0.004	−0.001	0.000	−0.001	−0.002	−0.003						
	t-stat	11.279***	4.007***	8.823***	10.340***	−0.984	−0.602	−6.783***	−3.057***	0.031	−1.158	−5.802***	−8.310***						
Japan	Mean	0.599	0.347	0.396	0.436	0.562	0.464	0.554	0.184	0.226	0.462	0.217	0.203	−0.045	−0.163	−0.171	0.025	−0.344	−0.261
	t-stat	7.968***	4.159***	11.490***	6.483***	10.143***	4.582***	9.877***	4.248***	5.750***	14.462***	6.230***	2.148**	−1.879	−3.609	−3.309	0.626	−14.850	−10.212
	Slope	−0.005	−0.005	0.000	−0.004	−0.003	−0.006	−0.003	−0.002	−0.002	−0.002	−0.002	−0.005						
	t-stat	−22.501***	−20.774***	−0.334	−16.925***	−5.845***	−30.124***	−4.292***	−5.374***	−2.626***	−3.674***	−7.198***	−5.403***						
Switzerland	Mean	0.379	0.303	−0.002	0.262	0.060	0.057	0.766	0.337	0.240	0.609	0.078	0.247	0.387	0.034	0.242	0.348	0.018	0.190
	t-stat	6.208***	2.838***	−0.100	3.035***	1.945*	2.081**	85.503***	31.262***	7.887***	7.435***	2.171**	9.428***	5.630***	0.345	6.346***	2.075**	2.343**	3.794***
	Slope	0.004	0.006	−0.001	0.005	0.002	0.002	−0.001	0.001	−0.001	−0.004	0.002	−0.001						
	t-stat	7.627***	7.542***	−1.895*	4.118***	2.733***	6.874***	−5.385***	1.916*	−0.675	−3.289***	3.732***	−1.596						
U.K.	Mean	0.610	0.418	0.372	0.416	0.289	0.181	0.865	0.420	0.488	0.706	0.144	0.253	0.255	0.002	0.116	0.290	−0.145	0.072
	t-stat	11.103***	8.147***	3.828***	4.013***	6.976***	2.299**	103.958***	11.707***	8.720***	108.395***	3.148***	9.983***	4.169***	0.022	0.765	2.723***	−1.846	0.760
	Slope	0.003	0.003	0.005	0.006	0.002	0.004	0.000	−0.001	−0.003	0.000	0.000	0.000						
	t-stat	2.763***	6.044***	4.119***	6.290***	3.036***	5.519***	−1.020	−1.277	−4.016***	−0.170	−0.303	−0.474						
U.S.	Mean	0.788	0.760	0.863	0.795	0.737	0.855	0.938	0.586	0.877	0.932	0.806	0.749	0.150	−0.174	0.014	0.136	0.069	−0.106
	t-stat	9.409***	7.895***	16.340***	10.312***	6.350***	24.099***	149.022***	9.558***	92.660***	57.456***	99.408***	78.787***	1.704**	−3.532	0.227	1.466*	0.609	−2.468
	Slope	0.005	0.006	0.003	0.004	0.007	0.002	0.000	0.003	0.000	−0.001	0.000	0.000						
	t-stat	7.612***	8.823***	8.370***	10.311***	8.403***	16.424***	−2.596**	5.268***	−2.167**	−4.973***	0.905	−0.933						

Table 1.3: Beta exposures of the local factors to the global factors.

For each sample country, each of the six local factors (MKT, SMB, HML, MOM, RMW, CMA) is regressed on the global factors over five-year rolling windows for the whole sample period of July 1990 – December 2017. The global factors are constructed using the market value-weighted averages of the local factors of Australia, Canada, France, Germany, Hong Kong, Japan, Switzerland, the U.K., and the U.S. The time-series average beta exposures of each local factor to the global MKT, SMB, HML, MOM, RMW, and CMA factors are reported. The proportions of significant beta exposures (5% level of significance) from all five-year rolling windows over the study period are presented.

	Local	Beta exposures						Proportion of significance					
Country	Factor	GMKT	GSMB	GHML	GMOM	GRMW	GCMA	GMKT	GSMB	GHML	GMOM	GRMW	GCMA
Australia	MKT	1.10	0.04	0.28	0.05	−0.24	−0.55	100.00%	19.56%	18.82%	6.27%	27.31%	22.88%
	SMB	−0.03	0.44	0.14	−0.06	−0.23	−0.20	0.37%	23.62%	1.48%	1.48%	0.00%	5.54%
	HML	0.06	−0.13	0.28	−0.13	0.10	0.25	17.71%	9.96%	19.93%	19.19%	15.50%	23.99%
	MOM	0.11	0.05	0.04	0.56	0.42	0.09	6.64%	6.27%	0.00%	77.49%	11.81%	0.00%
	RMW	−0.17	−0.27	−0.07	0.02	0.74	0.57	16.61%	23.99%	1.85%	5.54%	25.83%	25.09%
	CMA	−0.02	−0.09	−0.06	−0.04	0.07	0.69	9.96%	12.55%	3.69%	18.45%	28.41%	48.34%
Canada	MKT	0.97	0.14	0.58	0.08	−0.51	−0.89	100.00%	3.32%	52.77%	36.16%	41.33%	73.43%
	SMB	−0.03	0.74	−0.26	0.01	−0.41	0.22	11.81%	76.75%	12.18%	10.33%	43.91%	6.27%
	HML	0.01	0.00	0.55	−0.20	0.45	0.21	2.95%	1.85%	37.27%	13.28%	18.45%	21.40%
	MOM	0.17	−0.08	0.33	1.07	0.58	−0.20	17.71%	4.06%	18.82%	97.79%	41.33%	0.00%
	RMW	−0.10	−0.44	0.22	0.06	1.67	0.64	25.46%	26.20%	2.21%	14.39%	67.16%	22.14%
	CMA	−0.06	−0.05	−0.27	−0.05	0.50	1.07	5.17%	9.59%	39.48%	9.59%	46.86%	57.93%
France	MKT	1.18	−0.21	0.24	0.04	0.04	−0.01	100.00%	10.70%	14.39%	5.17%	0.00%	0.37%
	SMB	−0.33	0.81	0.10	−0.08	0.14	−0.03	74.91%	98.15%	5.54%	2.95%	2.95%	0.74%
	HML	0.03	0.15	0.56	−0.10	−0.13	−0.19	20.66%	8.86%	46.13%	30.63%	29.15%	32.10%
	MOM	−0.09	0.14	−0.07	0.69	0.33	−0.30	14.76%	5.54%	12.92%	86.35%	16.24%	11.44%
	RMW	0.10	0.11	0.11	0.06	0.36	0.13	14.39%	31.73%	12.92%	1.48%	23.99%	16.24%
	CMA	0.06	0.05	0.00	0.02	0.23	0.60	21.40%	0.00%	7.38%	10.33%	17.71%	50.92%
Germany	MKT	1.15	−0.10	0.09	0.08	0.21	0.10	100.00%	4.06%	7.01%	11.44%	1.11%	0.00%
	SMB	−0.32	0.58	−0.12	−0.15	0.20	0.17	71.22%	62.73%	7.01%	31.00%	11.44%	23.99%
	HML	0.10	0.09	0.54	0.01	−0.14	0.18	1.11%	6.27%	41.70%	4.06%	0.00%	8.49%
	MOM	−0.14	0.04	0.07	0.74	−0.10	0.02	20.66%	4.43%	7.38%	82.66%	2.21%	1.85%
	RMW	0.02	0.20	0.29	0.00	0.41	−0.09	11.81%	25.09%	26.57%	2.58%	28.04%	17.34%
	CMA	−0.18	0.16	−0.28	−0.08	−0.06	0.63	25.83%	5.17%	15.50%	10.33%	9.96%	28.41%

Country	Local Factor	Beta exposures GMKT	GSMB	GHML	GMOM	GRMW	GCMA	Proportion of significance GMKT	GSMB	GHML	GMOM	GRMW	GCMA
Hong Kong	MKT	1.17	−0.04	0.05	−0.08	−0.03	−0.51	100.00%	14.39%	21.40%	2.95%	0.37%	18.08%
	SMB	−0.20	1.04	0.46	0.14	−0.26	−0.40	1.11%	59.78%	4.43%	4.06%	11.81%	0.00%
	HML	0.05	0.02	0.24	−0.04	−0.03	0.26	0.37%	13.65%	3.32%	0.00%	1.48%	0.37%
	MOM	−0.23	0.00	−0.01	0.56	0.17	−0.18	14.02%	1.11%	16.24%	51.66%	0.00%	0.00%
	RMW	−0.07	0.30	0.03	−0.05	0.61	0.44	6.64%	7.01%	4.80%	15.13%	8.49%	0.00%
	CMA	−0.03	0.16	0.09	−0.02	0.12	0.23	0.37%	0.37%	0.74%	24.35%	0.00%	17.34%
Japan	MKT	0.95	0.13	−0.15	−0.08	−0.10	0.46	100.00%	8.12%	11.07%	13.65%	0.37%	21.03%
	SMB	−0.05	0.77	−0.09	−0.03	0.01	0.09	1.85%	89.67%	2.95%	11.07%	2.21%	1.11%
	HML	−0.08	0.04	0.84	0.00	−0.33	−0.29	24.72%	0.37%	83.03%	11.07%	6.27%	21.77%
	MOM	−0.04	−0.17	0.30	0.85	−0.09	−0.24	10.70%	0.37%	9.59%	98.52%	0.00%	13.28%
	RMW	0.05	−0.09	0.09	0.08	0.56	−0.42	26.57%	26.20%	17.34%	18.82%	42.80%	45.02%
	CMA	0.01	0.08	−0.18	−0.08	0.00	0.88	0.37%	5.90%	25.09%	19.93%	0.00%	79.34%
Switzerland	MKT	0.95	−0.09	0.20	0.06	0.27	0.04	100.00%	9.23%	4.06%	0.00%	9.59%	2.21%
	SMB	−0.23	0.63	0.13	−0.06	−0.10	−0.21	43.54%	73.80%	1.11%	20.30%	14.76%	14.39%
	HML	−0.01	−0.04	0.41	0.09	−0.28	0.15	5.17%	1.48%	26.20%	12.18%	12.55%	5.17%
	MOM	−0.15	−0.03	−0.33	0.67	−0.14	0.28	33.58%	5.90%	16.97%	80.07%	3.69%	16.61%
	RMW	0.12	0.15	0.06	−0.03	0.43	0.12	9.23%	12.18%	5.90%	5.54%	21.77%	17.34%
	CMA	−0.01	−0.24	−0.25	−0.09	0.05	1.01	8.12%	31.37%	15.50%	20.66%	0.00%	53.87%
U.K.	MKT	0.98	−0.19	0.19	−0.05	0.05	−0.15	100.00%	20.66%	9.96%	0.37%	7.38%	8.86%
	SMB	−0.14	0.92	0.07	−0.04	0.06	−0.01	28.04%	97.42%	27.31%	1.48%	0.37%	7.01%
	HML	0.05	−0.09	0.86	−0.12	−0.17	−0.33	16.61%	3.69%	81.55%	20.30%	12.18%	12.18%
	MOM	−0.06	−0.04	−0.35	0.82	0.17	0.32	0.00%	9.59%	15.13%	95.20%	1.85%	7.38%
	RMW	0.01	0.19	−0.21	0.05	0.64	−0.14	24.72%	41.70%	35.06%	4.43%	58.30%	18.45%
	CMA	0.06	−0.03	0.07	−0.02	−0.07	0.76	5.90%	7.38%	12.55%	9.96%	2.58%	61.25%
U.S.	MKT	0.89	0.00	−0.07	0.01	0.03	−0.13	100.00%	25.09%	10.70%	2.58%	2.95%	28.04%
	SMB	0.14	1.17	−0.03	0.04	0.08	0.02	52.03%	100.00%	2.21%	9.96%	0.00%	15.13%
	HML	−0.01	−0.04	1.33	0.04	0.10	0.08	10.70%	1.48%	100.00%	7.75%	9.23%	41.70%
	MOM	0.07	0.05	−0.07	1.16	−0.09	0.09	30.26%	2.58%	12.18%	100.00%	7.38%	17.71%
	RMW	0.00	−0.03	−0.07	−0.03	1.25	0.04	14.02%	12.92%	30.63%	16.97%	100.00%	0.74%
	CMA	0.01	−0.02	0.15	0.03	−0.05	1.13	5.54%	45.02%	37.64%	16.61%	28.04%	100.00%

Table 1.4: Performance comparisons of the local, global, and pure local factor models: Evidence from individual stocks.

The individual stock excess returns in each sample country are regressed on the local, global, and pure local factors, respectively, over five-year rolling windows from July 1990 to December 2017. The pure local factors are the residuals from the regressions of the local factors on the global factors. A stock is included in the regression only if the stock has all 60 monthly returns during the five-year period. In Panel A, the differences between the time series means of the average $|\alpha|$ (in percent) and average R^2 from the regressions of individual stock excess returns on the local and global factors are presented. In addition, we formally test whether the differences in the average $|\alpha|$s and R^2s diminish over time. Considering the serial correlations in the residuals from the time-trend regressions, the Cochrane-Orcutt adjusted slopes and the associated t-statistics (adjusted t-stat) are reported. In Panel B, the time-series means and standard deviations of the cross-sectional average $|\alpha|$s and R^2s from the regressions of individual stock excess returns on the local, global, and pure local factors are presented.

Panel A: Differences of $|\alpha|$s and R^2s between the local and global factor models

	Difference in $|\alpha|$			Difference in R^2		
		Time trends			Time trends	
	Mean	Adjusted slope	Adjusted t-stat	Mean	Adjusted slope	Adjusted t-stat
Australia	-0.20	-0.0010	-6.07	0.14	0.0004	0.23
Canada	-0.10	-0.0052	-2.13	0.05	0.0018	6.17
France	-0.12	0.0008	2.11	0.10	-0.0006	-10.34
Germany	-0.22	0.0008	6.80	0.10	-0.0009	-11.40
Hong Kong	-0.33	-0.0005	-4.21	0.20	-0.0007	-5.60
Japan	-0.29	0.0010	2.84	0.17	-0.0018	-15.41
Switzerland	-0.16	-0.0008	-27.65	0.12	-0.0002	-11.60
U.K.	-0.03	0.0004	1.12	0.08	0.0008	12.07
U.S.	-0.12	-0.0004	-11.67	0.03	0.0001	10.02

Panel B: Individual firm $|\alpha|$ and R^2 from alternative factor models

		Local		Global		Pure Local	
		$|\alpha|$	R^2	$|\alpha|$	R^2	$|\alpha|$	R^2
Australia	Mean	2.03	0.27	2.24	0.17	2.28	0.08
	Std dev	0.27	0.02	0.26	0.03	0.35	0.03
Canada	Mean	2.50	0.19	2.65	0.14	2.74	0.04
	Std dev	0.16	0.03	0.07	0.04	0.20	0.01
France	Mean	1.39	0.27	1.49	0.19	1.69	0.06
	Std dev	0.14	0.03	0.11	0.04	0.16	0.03
Germany	Mean	1.65	0.23	1.75	0.17	1.88	0.05
	Std dev	0.29	0.02	0.18	0.04	0.28	0.04
Hong Kong	Mean	1.94	0.28	2.34	0.13	2.19	0.13
	Std dev	0.17	0.02	0.17	0.04	0.21	0.05
Japan	Mean	1.11	0.30	1.35	0.13	1.45	0.15
	Std dev	0.11	0.07	0.08	0.09	0.07	0.03
Switzerland	Mean	1.10	0.31	1.30	0.20	1.40	0.10
	Std dev	0.05	0.03	0.07	0.04	0.06	0.02
U.K.	Mean	1.67	0.21	1.71	0.15	1.86	0.05
	Std dev	0.11	0.03	0.11	0.02	0.12	0.02
U.S.	Mean	2.16	0.18	2.28	0.16	2.40	0.02
	Std dev	0.11	0.02	0.08	0.04	0.13	0.01
Average		1.73	0.25	1.90	0.16	1.99	0.08
		0.47	0.05	0.49	0.02	0.45	0.04

Table 1.5: Performance comparisons of the local, global, and pure local factor models: Evidence from portfolios.

For Canada, Japan, the U.K., and the U.S., 25 size and book-to-market portfolios (size-BE/ME) are constructed using 5×5 independent sorts on size and BE/ME each year. Similarly, 25 size and momentum portfolios (size-MOM) are constructed using 5×5 independent sorts on size and MOM each year. For each country, the dollar-denominated returns on each portfolio in excess of the U.S. risk-free rate are regressed on the local, global, and pure local factors over the period of July 1990 – December 2017. In Panel A, the cross-sectional averages of the $|\alpha|$s and R^2s from the regressions of the portfolio excess returns on the local, global, and pure local factors are reported for each sample country, with the standard deviations in parentheses. In Panel B, the Gibbons, Ross, and Shanken (1989) F-statistic (F-stat) and p-value (p-val) for the size and book-to-market and the size and momentum portfolios are reported for each sample country.

Panel A: Averages of $|\alpha|$s and R^2s from portfolio regressions

		Local		Global		Pure Local	
		$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2
Canada	Size-BE/ME	0.39	0.63	0.31	0.48	0.81	0.26
		(0.45)	(0.14)	(0.36)	(0.12)	(0.50)	(0.06)
	Size-MOM	0.65	0.62	0.66	0.42	1.06	0.24
		(0.56)	(0.13)	(0.58)	(0.12)	(0.71)	(0.05)
Japan	Size-BE/ME	0.29	0.75	0.59	0.38	1.03	0.40
		(0.14)	(0.07)	(0.22)	(0.07)	(0.20)	(0.06)
	Size-MOM	0.35	0.75	0.52	0.38	1.05	0.40
		(0.28)	(0.05)	(0.22)	(0.04)	(0.39)	(0.06)
U.K.	Size-BE/ME	0.15	0.78	0.23	0.50	0.64	0.27
		(0.10)	(0.10)	(0.14)	(0.10)	(0.16)	(0.05)
	Size-MOM	0.33	0.79	0.37	0.52	0.64	0.25
		(0.28)	(0.07)	(0.37)	(0.10)	(0.44)	(0.07)
U.S.	Size-BE/ME	0.10	0.89	0.38	0.71	0.58	0.16
		(0.07)	(0.04)	(0.14)	(0.05)	(0.21)	(0.04)
	Size-MOM	0.09	0.88	0.50	0.72	0.56	0.14
		(0.11)	(0.03)	(0.22)	(0.04)	(0.33)	(0.04)

Panel B: GRS test results

		F-stat	p-val	F-stat	p-val	F-stat	p-val
Canada	Size-BE/ME	1.01	0.45	0.85	0.67	1.31	0.15
	Size-MOM	3.12	0.00	3.21	0.00	4.16	0.00
Japan	Size-BE/ME	1.39	0.11	1.30	0.16	1.87	0.01
	Size-MOM	3.11	0.00	2.39	0.00	4.76	0.00
U.K.	Size-BE/ME	1.28	0.17	1.25	0.19	1.39	0.11
	Size-MOM	4.97	0.00	5.72	0.00	6.07	0.00
U.S.	Size-BE/ME	1.73	0.02	2.82	0.00	3.08	0.00
	Size-MOM	1.57	0.04	2.37	0.00	2.99	0.00

Table 1.6: Performance comparisons of the local, global, and pure local factor models: Evidence from nine countries.

For Australia, Canada, France, Germany, Hong Kong, Japan, Switzerland, the U.K. and the U.S., nine (3×3) size and book-to-market portfolios and nine (3×3) size and momentum portfolios are constructed using independent sorts. For each country, the dollar-denominated returns on each portfolio in excess of the U.S. risk-free rate are regressed on the local, global, and pure local factors over the period of January 1996 – December 2017. The cross-sectional average $|\alpha|$s and R^2s and Gibbons, Ross, and Shanken (1989) F-statistic (F-stat) and p-value (p-val) are reported.

	Local		Global		Pure Local		Local		Global		Pure Local		Local		Global		Pure Local	
	Australia						Canada						France					
	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2
Size-BM/ME	0.29	0.79	0.33	0.53	1.28	0.25	0.39	0.77	0.32	0.58	1.18	0.19	0.27	0.86	0.18	0.58	0.92	0.27
Size-MOM	0.47	0.81	0.60	0.54	1.14	0.26	0.46	0.77	0.51	0.54	1.06	0.22	0.36	0.87	0.30	0.60	0.94	0.25
GRS test	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val
Size-BM/ME	2.01	0.04	2.05	0.03	2.81	0.00	1.68	0.09	0.83	0.59	2.64	0.01	4.04	0.00	1.66	0.10	5.09	0.00
Size-MOM	6.92	0.00	9.80	0.00	11.75	0.00	3.46	0.00	3.65	0.00	6.50	0.00	5.55	0.00	4.71	0.00	8.90	0.00
	Germany						Hong Kong						Japan					
	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2
Size-BM/ME	0.21	0.84	0.12	0.56	0.81	0.27	0.28	0.76	0.42	0.36	0.88	0.40	0.20	0.90	0.28	0.28	0.82	0.61
Size-MOM	0.40	0.84	0.31	0.57	0.82	0.25	0.49	0.75	0.67	0.35	1.09	0.40	0.25	0.90	0.29	0.30	0.83	0.60
GRS test	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val
Size-BM/ME	2.68	0.01	0.98	0.46	4.13	0.00	3.24	0.00	4.17	0.00	4.91	0.00	2.23	0.02	2.14	0.03	3.10	0.00
Size-MOM	8.92	0.00	5.94	0.00	11.56	0.00	4.24	0.00	5.43	0.00	5.73	0.00	2.58	0.01	2.54	0.01	5.82	0.00
	Switzerland						U.K.						U.S.					
	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2
Size-BM/ME	0.18	0.79	0.20	0.51	0.86	0.28	0.12	0.87	0.16	0.58	0.65	0.27	0.06	0.95	0.12	0.85	0.86	0.08
Size-MOM	0.28	0.81	0.28	0.53	0.80	0.27	0.31	0.87	0.30	0.61	0.69	0.24	0.08	0.93	0.25	0.85	0.80	0.06
GRS test	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val
Size-BM/ME	1.37	0.20	0.64	0.76	2.60	0.01	1.19	0.30	1.01	0.44	2.20	0.02	0.81	0.61	3.44	0.00	3.72	0.00
Size-MOM	3.14	0.00	3.37	0.00	5.71	0.00	8.15	0.00	9.20	0.00	10.33	0.00	1.50	0.15	3.55	0.00	4.56	0.00

Table 1.7: Performance comparisons of the local, global, and pure local factor models: Evidence from the size and R^2 sorted portfolios. For the U.S., Japan, the U.K. and Canada, 25 (5×5) size and R^2 portfolios are constructed through independent sorts on size and R^2s each month from July 1995 to December 2017. Each month, we first sort all individual stocks based on the previous month size and the R^2s estimated from the regressions of individual stock excess returns on the global factors over the previous five-year period. Then, 25 size and R^2 sorted portfolios are formed every month using the value-weighted stock returns within each portfolio. This table reports the absolute intercepts ($|\alpha|$s) and the R^2s from regressions of the 25 size-R^2 sorted portfolio excess returns on the local, global, and pure local factors, respectively, and the Gibbons, Ross, and Shanken (1989) F-statistic of the portfolios, with the p-values in parentheses.

	Local										Global										Pure Local									
	$\|\alpha\|$					R^2					$\|\alpha\|$					R^2					$\|\alpha\|$					R^2				
	Seg	2	3	4	Int	Seg	2	3	4	Int	Seg	2	3	4	Int	Seg	2	3	4	Int	Seg	2	3	4	Int	Seg	2	3	4	Int
Canada																														
Small	1.57	1.53	2.00	2.04	2.05	0.51	0.50	0.55	0.44	0.33	1.41	1.68	1.75	2.00	2.19	0.32	0.31	0.33	0.30	0.21	2.32	2.19	2.65	2.46	2.20	0.19	0.19	0.23	0.13	0.13
2	0.39	0.31	0.32	0.21	0.57	0.55	0.67	0.62	0.60	0.48	0.44	0.19	0.17	0.09	0.78	0.33	0.44	0.38	0.37	0.38	0.93	1.00	0.98	0.87	1.23	0.21	0.22	0.22	0.23	0.12
3	0.17	0.19	0.05	0.18	0.08	0.60	0.65	0.73	0.68	0.66	0.11	0.00	0.02	0.05	0.06	0.50	0.49	0.54	0.53	0.53	0.54	0.87	0.71	0.78	0.45	0.10	0.17	0.20	0.16	0.14
4	0.52	0.10	0.51	0.50	0.08	0.59	0.63	0.74	0.76	0.76	0.33	0.06	0.19	0.31	0.20	0.45	0.53	0.56	0.58	0.66	1.14	0.79	1.10	1.17	0.64	0.15	0.11	0.19	0.18	0.12
Big	0.11	0.36	0.26	0.22	0.08	0.32	0.44	0.50	0.66	0.86	0.26	0.08	0.08	0.06	0.29	0.18	0.31	0.31	0.54	0.73	0.37	0.86	0.75	0.78	0.85	0.14	0.16	0.20	0.13	0.15
GRS stat	2.79	(0.00)									2.43	(0.00)									3.35	(0.00)								
Japan																														
Small	0.46	0.56	0.90	0.85	0.73	0.86	0.88	0.89	0.88	0.84	0.05	0.13	0.36	0.37	0.16	0.22	0.25	0.24	0.22	0.20	1.08	1.25	1.55	1.58	1.40	0.63	0.63	0.65	0.66	0.64
2	0.15	0.09	0.32	0.35	0.49	0.87	0.91	0.90	0.91	0.88	0.54	0.34	0.14	0.14	0.00	0.21	0.27	0.28	0.25	0.23	0.47	0.71	0.94	1.00	1.12	0.65	0.64	0.62	0.65	0.65
3	0.01	0.13	0.23	0.16	0.25	0.87	0.89	0.90	0.89	0.87	0.37	0.25	0.17	0.26	0.19	0.22	0.25	0.25	0.25	0.27	0.60	0.74	0.83	0.80	0.87	0.65	0.63	0.64	0.63	0.60
4	0.00	0.04	0.13	0.04	0.07	0.81	0.84	0.85	0.88	0.84	0.19	0.15	0.19	0.32	0.36	0.18	0.20	0.26	0.28	0.33	0.53	0.60	0.69	0.64	0.66	0.62	0.63	0.59	0.59	0.52
Big	0.32	0.20	0.28	0.09	0.04	0.71	0.75	0.79	0.87	0.86	0.01	0.16	0.23	0.44	0.58	0.22	0.28	0.34	0.43	0.52	0.81	0.65	0.75	0.58	0.39	0.49	0.48	0.45	0.43	0.36
GRS stat	3.40	(0.00)									2.75	(0.00)									3.62	(0.00)								
U.K.																														
Small	0.11	0.23	0.13	0.09	0.91	0.70	0.71	0.63	0.43	0.24	0.13	0.11	0.25	0.18	0.62	0.42	0.38	0.35	0.27	0.11	0.35	0.23	0.45	0.32	0.53	0.27	0.33	0.26	0.15	0.13
2	0.16	0.03	0.00	0.22	0.05	0.78	0.81	0.81	0.75	0.69	0.04	0.14	0.13	0.31	0.12	0.42	0.47	0.48	0.42	0.42	0.36	0.51	0.51	0.69	0.46	0.33	0.32	0.32	0.31	0.27
3	0.05	0.06	0.16	0.16	0.35	0.72	0.79	0.84	0.83	0.75	0.11	0.13	0.21	0.15	0.20	0.43	0.51	0.54	0.56	0.58	0.54	0.60	0.69	0.73	0.90	0.27	0.26	0.28	0.25	0.16
4	0.20	0.07	0.17	0.36	0.30	0.51	0.69	0.80	0.84	0.82	0.24	0.08	0.07	0.28	0.13	0.34	0.45	0.61	0.64	0.70	0.59	0.60	0.76	0.92	0.91	0.15	0.25	0.18	0.19	0.13
Big	0.30	0.12	0.05	0.01	0.05	0.17	0.51	0.74	0.82	0.93	0.22	0.15	0.13	0.12	0.12	0.12	0.35	0.53	0.65	0.79	0.74	0.52	0.53	0.52	0.54	0.06	0.16	0.21	0.17	0.13
GRS stat	0.85	(0.67)									0.51	(0.98)									0.95	(0.54)								
U.S.																														
Small	0.32	0.41	0.87	0.59	0.75	0.54	0.55	0.59	0.60	0.59	0.29	0.41	0.89	0.68	1.06	0.62	0.60	0.63	0.61	0.58	0.51	0.43	0.87	0.37	0.36	0.00	0.01	0.00	0.02	0.02
2	0.02	0.22	0.26	0.28	0.59	0.71	0.78	0.81	0.83	0.83	0.03	0.19	0.30	0.38	0.93	0.70	0.76	0.76	0.76	0.76	0.59	0.83	0.81	0.80	0.91	0.02	0.02	0.04	0.05	0.05
3	0.01	0.11	0.21	0.19	0.37	0.82	0.87	0.90	0.94	0.92	0.00	0.11	0.24	0.29	0.70	0.70	0.74	0.79	0.82	0.82	0.88	1.06	1.11	1.07	1.11	0.10	0.11	0.09	0.09	0.07
4	0.05	0.00	0.02	0.02	0.17	0.83	0.88	0.91	0.95	0.94	0.05	0.06	0.09	0.16	0.51	0.72	0.76	0.78	0.84	0.84	0.88	1.05	1.03	0.97	1.06	0.10	0.10	0.11	0.09	0.07
Big	0.14	0.10	0.11	0.05	0.00	0.52	0.76	0.82	0.90	0.96	0.20	0.08	0.07	0.17	0.34	0.49	0.65	0.74	0.83	0.90	0.72	0.58	0.44	0.62	0.53	0.02	0.09	0.04	0.04	0.04
GRS stat	1.30	(0.16)									1.52	(0.06)									1.68	(0.03)								

Table 1.8: Replication of the Griffin study.

For Canada, Japan, the U.K., and the U.S., the individual stock excess returns and 25 size and book-to-market portfolio excess returns are regressed on the local, global, international, and pure local factors over five-year rolling windows from January 1981 to December 1995. Following Griffin (2002), different versions of the Fama and French three factors (MKT, SMB, HML) are considered in each regression model. In Panel A, the regression results of the individual stocks are reported. In Panel B, the regression results of the 25 size and book-to-market portfolios are reported. In Panel C, the Gibbons, Ross, and Shanken (1989) F-statistic (F-stat) and p-value (p-val) for the size and book-to-market portfolios are reported.

	Weighted factor regressions								Unweighted factor regressions							
	Local		Global		International		Pure Local		Local		Global		International		Pure Local	
	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2	$\|\alpha\|$	R^2
Panel A: Average $\|\alpha\|$ and R^2 of individual firms																
Canada																
Mean	1.46	0.18	1.58	0.09	1.53	0.19	1.78	0.09	1.45	0.18	1.63	0.11	1.55	0.19	1.73	0.07
Std dev	(0.07)	(0.01)	(0.04)	(0.02)	(0.07)	(0.02)	(0.07)	(0.00)	(0.07)	(0.02)	(0.08)	(0.02)	(0.09)	(0.02)	(0.08)	(0.00)
Japan																
Mean	0.94	0.43	1.30	0.27	1.02	0.43	1.53	0.14	0.87	0.44	1.28	0.27	0.95	0.45	1.32	0.15
Std dev	(0.08)	(0.07)	(0.24)	(0.07)	(0.07)	(0.07)	(0.28)	(0.02)	(0.06)	(0.07)	(0.26)	(0.05)	(0.08)	(0.06)	(0.31)	(0.01)
U.K.																
Mean	1.16	0.31	1.29	0.15	1.30	0.31	1.75	0.15	1.16	0.31	1.30	0.20	1.29	0.31	1.81	0.10
Std dev	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.02)	(0.27)	(0.01)	(0.02)	(0.02)	(0.06)	(0.02)	(0.05)	(0.02)	(0.30)	(0.00)
U.S.																
Mean	1.23	0.21	1.48	0.14	1.42	0.21	1.51	0.07	1.24	0.21	1.53	0.11	1.43	0.22	1.44	0.10
Std dev	(0.08)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.05)	(0.02)	(0.08)	(0.01)	(0.04)	(0.01)	(0.03)	(0.01)	(0.06)	(0.01)
Panel B: Average $\|\alpha\|$ and R^2 of size and book-to-market portfolios																
Canada	0.38	0.58	0.45	0.29	0.42	0.59	1.00	0.26	0.38	0.59	0.42	0.34	0.37	0.59	0.97	0.21
Japan	0.19	0.81	0.29	0.44	0.22	0.82	1.32	0.37	0.19	0.87	0.21	0.48	0.17	0.87	0.96	0.37
U.K.	0.20	0.80	0.40	0.39	0.26	0.80	1.05	0.41	0.25	0.83	0.37	0.55	0.28	0.83	0.96	0.27
U.S.	0.18	0.90	0.25	0.62	0.19	0.90	0.59	0.25	0.17	0.92	0.42	0.48	0.19	0.92	0.51	0.42
	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val	F-stat	p-val
Panel C: GRS test results of size and book-to-market portfolios																
Canada	1.59	0.05	1.59	0.05	1.52	0.07	2.24	0.00	1.56	0.06	1.47	0.09	1.44	0.10	2.29	0.00
Japan	0.64	0.90	0.99	0.49	1.00	0.48	1.63	0.04	0.65	0.90	0.69	0.86	0.65	0.89	1.73	0.02
U.K.	1.29	0.18	1.49	0.08	1.33	0.15	2.19	0.00	1.42	0.11	1.15	0.30	1.19	0.26	2.09	0.00
U.S.	4.06	0.00	4.16	0.00	4.18	0.00	5.99	0.00	4.06	0.00	4.61	0.00	4.28	0.00	6.13	0.00

CHAPTER 2

DYNAMIC FIRM-SPECIFIC NETWORK AND RISK DIVERSIFICATION

2.1 Introduction

Firms are intricately linked to each other through many types of relationships. Some of these links are clear and contractual, such as the customer-supplier links and the trading activities between firms, while others are implicit and less transparent, such as the cultures and political strategies of firms.[1] These links do not exist as independent relationships, but interact in a complex way resulting in the network of firms, which is affected by both the firms and the market. A good number of studies examined the applications of one or a few links, such as the information flow and the return predictability across the economically linked firms. Some other studies proposed alternative ways to measure the network, including both the pairwise and system-wide approaches.[2] However, to the best of our knowledge, none of the studies distinguishes the firm-specific effects on the network from other firms and the market, and as a result, there is no study on the measurement of the firm-specific network and its implications.

However, there are many reasons to investigate the firm-specific network. First, it assists in identifying the periods of high systemic risk, then in further understanding and measuring the risk. During the periods of high systemic risk, financial institutions are closely interrelated and likely to fail together, thus the firm-specific relationships are weak. Illiquidity, insolvency, and losses among the interconnected institutions can quickly propagate during the periods of financial distress, directly through the firm-specific linked firms or indirectly through other firms or the market. Therefore, the downward time in the number of the firm-specific relations indicates the possible periods of high systemic risk and the firm-specific relationships help uncover the channels of risk transmission, which provide valuable information for the design of financial regulations. Second, the firm-specific network can be applied to examine the existence of unknown common risk factors omitted in existing factor models. The firm-specific network describes the firm-specific relation-

[1] See, for example, [31], [32], [33], and [34].
[2] See, for example, [31], [35], and [36].

ships after the common information conveyed from other firms and the market are removed. Thus, a set consisted of all common risk factors should capture all these common information indicated by the firm-specific relationships, otherwise it is not the full set of risk factors. Third, investors can consistently gain from risk diversification and a low transaction cost by holding a few stocks. The dynamic firm-specific network describes the dependent and independent relationships between the firms conditional on other firms or the market over time. In principle, if investors can identify the conditional independent and dependent stocks, they can hold few independent stocks to adequately diversify the idiosyncratic risk. Therefore, they do not need to bear the idiosyncratic risk due to the restrictions of investment (e.g., participates in employee retirement plans) and the generally increasing correlations among stocks, especially during crises. Certainly, there are many more other important implications of the firm-specific network than just these.

In this paper, we propose a system-wide approach to estimate the firm-specific network and then examine the implications of the estimated network. A system-wide approach to measure the dynamic firm-specific relationships is essential to the study of the firm-specific network, because neglecting the system-wide impact on the estimation of the firm-specific relationships could result in spurious inferences.[3] Specifically, we (i) estimate the firm-specific relationships conditional on all other firms in the market per year over the sample period from 1980 to 2015, (ii) examine whether the firm attributes affect the firm-specific relationships, if so, then how large the effects are and how the effects vary over time, (iii) utilize the estimated firm-specific network to explore the existence of unknown common risk factors, and (iv) apply the estimated firm-specific network to consistently gain from risk diversification. In doing so, we first estimate the firm-specific network by the graph estimation method proposed by [37] for a sample of 1,000 largest individual firms per year, which approximately represent the market, comprising 92% of the total market value on average over the sample period of 1980–2015. Then, the firm-specific relationships are summarized as 0 or 1 each year, representing the conditional independence or dependence given the entire market, respectively.

The findings regarding the firm-specific network show that both the aggregate and individual degrees of the firm-specific connections decrease over the sample period of 1980–2015, especailly

[3]For example, [2] show a fundamental flaw in the most widely used pairwise correlation method, and [36] use the simulation method to show the incorrect inferences by a pairwise approach and the advantages of a system-wide approach.

during crisis years. In particular, the aggregate degree of the firm-specific connections decreases from 34,300 in 1980 to 25,500 in 2015 among the 1,000 sample firms. The median of the individual firm-specific connections reduces from 34 in 1980 to 25 in 2015. These downward time trends in the degrees of the firm-specific connections imply that the amount of information conveyed from the market increases over time, which suggests a new measurement of market integration. The less the aggregate degree of the firm-specific connections is, the more the market is integrated. By further comparing the new measurement with some popular measurements of integration, we find that the decreasing in the aggregate degree of the firm-specific connections leads to the increasing in the cross-sectional average pairwise correlation among firms and the increasing in the cross-sectional average degree of firm integration measured by R-square ([2]), significantly at 5% level, which points to the advantages of the new measurement of market integration — a system-wide and factor-free measurement.[4]

In addition, the apparent drops in the aggregate degree of the firm-specific connections during crisis years indicate the possible periods of high systemic risk. The analysis of change points formally demonstrates that the significant changes in the mean of the aggregate degree of the firm-specific connections happened all around crises, such as 1992 (Black Wednesday), 1999 (Asian/Russian financial crises), 2002 (IT bubble bust and 9/11 incident at 2001), 2007 (US subprime mortgage crisis), and 2010 (European sovereign debt crisis). At the meantime, the dynamic firm-specific relationships aid to identify the channels of risk transmission, either directly through the conditionally linked firms or indirectly through the market.

In order to look into what might affect the firm-specific relationships, we then apply the Bayesian nonparametric model, following [38], to explore the time-varying effects of the firm attributes on the firm-specific network. The results from the Bayesian nonparametric regression indicate that firms from the same industry or the same style defined by firms' market capitalization and book-to-market characteristics tend to associate with each other even after their common information from the market removed, whereas the geographical locations of firm headquarters do not contribute on the firm-specific relationships. Industry attributes have the greatest positive effect on the firm-

[4]See, for example, [2] show the pairwise correlation across markets is a poor measure of integration, and propose the R-square method to measure the degree of integration. The estimation of R-square depends on the factors used in the regression.

specific relationships, 1.197 on average over the sample period from 1980 to 2015, followed by style attributes, 0.0389 on average, then location attributes, 0.003 on average but insignificant. In particular, the industry effect increases more than twice over the sample period, going up sharply during the IT bubble bust and the U.S. subprime mortgage crisis, while the style effect has the same upward time trend, albeit increasing much slower.

Our prior findings showing the downward time trend in the aggregate degree of the firm-specific connections and the significant effects of the industry and style attributes on the firm-specific relationships have immediate implications in asset pricing and risk diversification. With the assumption that firm returns are multivariate Gaussian, the partial correlation among two firms is zero if and only if they are conditionally independent given the entire market ([39]). Based on this "if and only if" relationship, the comparison between the cross-sectional average partial correlation among firms given a set of existing known common risk factors and the firm conditional independence in the firm-specific network indicates the existence of unknown common factors besides the five factors of Fama and French ([7]) and the momentum factor ([8]). That is, the conditional relationships among firms in the firm-specific network provide a diagnostic criterion for the validity and completeness of a factor model. Moreover, this criterion does not depend on a factor structure in the model, such as a time-invariant factor effect (coefficient) or a time-varying factor effect.[5]

To further exploit the implications of the firm-specific network, we examine the potential gain from risk diversification using the conditional independent stocks. In principle, investors should be able to diversify more risk by holding a portfolio consisted of certain number of conditional independent stocks than holding a portfolio consisted of a comparable number of randomly selected stocks. That is, few independent stocks are needed to achieve a given level of diversification. In order to identify the groups of the conditional independent stocks, we apply the fast greedy community detection algorithm ([43]) to cluster the 1,000 sample firms into different communities based on the firm-specific network each year over the sample period from 1980 to 2015. Stocks within a community are more likely to be conditionally related, while stocks across communities are more likely to be conditionally independent. The results from the firm clustering show that the 1,000 sam-

[5]Empirical studies in asset pricing rely on linear multi-factor models with an assumption of either time-invariant or time-varying coefficients. Then, the time variation in the risk can bias the time-invariant estimates of alphas and betas, and therefore the conclusions of the asset pricing tests ([40], [41], and [42]).

ple firms are stably grouped into only four to six communities per year. In addition, the relations among firms within the same community become tight over time, particularly during crises, despite the deceasing number of the firm-specific connections.

Next, following [44], we construct an equally weighted portfolio with a varying number of stocks randomly chosen from each community with replacement, ranging from 1 to 20 stocks, and compute the portfolio variance using monthly stock returns. This procedure is repeated by 5,000 times. As can be expected, a few stocks are needed to achieve a given level of risk diversification when the stocks are chosen from different communities, because stocks within the same community share similar properties and are more likely to be correlated. Furthermore, when investors simply hold the equal-weighted stocks randomly selected from the communities, the gain from risk diversification as measured by a higher Sharpe ratio relative to the market index turns out to be consistent — not only ex-post and ex-ante, but also no matter the choices of stocks. In particular, the resulting equal-weighted portfolio, formed by at most three stocks randomly selected per community, yields an in-sample monthly Sharpe ratio of 0.169 on average (over the 5,000 replications) and wins the market index 93.3% times out of all 5,000 replications over the sample period of 1980–2015. When the number of stocks randomly selected per community increases to ten, the resulting portfolio yields an in-sample monthly Sharpe ratio of 0.181 on average and beats the market index 100% times out of all 5,000 replications without exception, and on the other hand, yields an out-of-sample monthly Sharpe ratio of 0.137 on average and wins the market index 69.4% times of all replications. It is noteworthy that this consistent gain is estimated simply using the equal-weighted portfolio, thus the gain can be considerably increased through mean and variance optimization strategy. Moreover, the gain here means not only the higher Sharpe ratio relative to the market index, but also the small number of stocks needed in risk diversification — a lower transaction cost.

In addition, the results from the studies of the resulting equal-weighted portfolios over the two evenly divided sub-sample periods of 1980–1997 and 1998–2015, demonstrate that the consistent gain of the resulting portfolio in risk diversification is much stronger over the later sub-sample period than the earlier sub-sample period. For example, when only one stock randomly selected per community per year, i.e., at most six stocks in total per year, the resulting equal-weighted portfolio yields an in-sample (out-of-sample) Sharpe ratio of 0.136 (0.103) on average and wins the market index 80.8% (64.7%) times out of all 5,000 replications over the later sub-sample period of 1998–2015. This stronger performance over the later sub-sample period is likely due to the increasing correlations among firms over time as the market is becoming more integrated, thus the risk diversification through the conditionally independent stocks becomes more valuable and beneficial.

Overall, the findings in this paper show that the firm-specific connections are important in the study of market integration, systemic risk, asset pricing, and investment. The downward time trend in the aggregate degree of the firm-specific connections suggests the possible periods of high systemic risk and offers a system-wide and factor-free measurement of market integration. The comparison between the firm-specific network and partial correlations given a set of known common risk factors indicates the existence of unknown risk factors, which provides a diagnostic criterion to the validity and completeness of existing factor models. Furthermore, the firm-specific network gives valuable insights into the gainful diversification. Investors can consistently benefit from risk diversification by holding few conditional independent stocks identified from the firm-specific connections, and the gain is becoming strong over time despite the increasing correlations among stocks over time.

The remainder of the paper is organized as follows. Section 2.2 describes the data and provides the estimation of the firm-specific relationships conditional on all other sample firms in the market. Section 2.3 establishes the effects of the firm attributes on the firm-specific relationships. Section 2.4 explores the existence of unknown common risk factors by the comparison of the estimated firm-specific network with the partial correlations given a set of known common risk factors. Section 2.5 details the implication of the firm-specific network in risk diversification. Lastly, Section 2.6 concludes this paper.

2.2 Dynamic Firm-specific Network

In this section, we focus on the identification of the firm-specific relationships given all other firms in the market. Consider a sample of p firms with returns $r_{1t_i}, r_{2t_i}, \cdots, r_{pt_i} \sim N_p(\mu_t, \Sigma_t)$ at year t, where $t_i = 1, \cdots, T$ and T is the number of firm's return observations within year t, which can be different across firms. [45] shows that firms i and j are conditional independent at year t given all other $p-2$ sample firms if and only if $\Sigma^{-1}_{ij,t} = 0$. In other words, when $\Sigma^{-1}_{ij,t} \neq 0$, firms i and j are significantly interconnected at year t, conditional on the information conveyed by the other $p-2$ sample firms. Therefore, we measure the firm-specific relationships by $\Sigma^{-1}_{ij,t}$ in this paper and empirically estimate the firm-specific network Σ^{-1}_t per year, allowing the firm-specific relationships vary over time.

2.2.1 Data description and sample selection

Each year over the whole sample period from 1980 to 2015, we use the daily stock returns in the U.S. within the year to estimate the firm-specific network Σ^{-1}_t, and then use the monthly stock returns for the inferences and applications of the estimated network.[6] The daily and monthly stock returns are obtained from the University of Chicago's Center for Research in Security Prices database (CRSP). Firm characteristics, such as common equity, total asset, global industry classification standard (GICS), and headquarter state location of firms, are obtained from Compustat.

We filter the stocks by the share code (SHRCD) equal to 10 or 11 to keep only the common shares, and the exchange code (EXCHCD) equal to 1, 2, or 3 to keep only stocks traded at NYSE, AMEX, or NASDAQ. We then require the stocks having all daily observations within the study year t and all monthly observations during both the current year t and the next year $t+1$ to ensure the estimation and inferences of Σ^{-1}_t.[7] Because the number of firms varies over time, in order to fairly compare the network each year and estimate the firm-specific network by removing the common information conveyed from the entire market, we employ only the largest 1,000 firms ($p = 1,000$) in the construction of the firm-specific network Σ^{-1}_t per year.[8]

Table 2.1 presents the summary statistics for the 1,000 sample firms. Each year, from 1980

[6]The firm-specific network is estimated each year, there is no preference on the starting year for the analyses. Looking back to 1980, there are 36 years in the sample period which assist the examination of what's going on in the firm-specific relationships over time.

[7]Note that the estimation method for Σ^{-1}_t proposed in Section 2.2.2 in this paper works for unbalanced panel data. Because the number of interrelationships among p firms, i.e., $p(p-1)$, to be identified is very large per year compared with the number of daily observations per year, we require the sample firms to have all daily observations within the study year to have more observations in the estimation.

[8]First, $p(p-1)$ interrelationships among p firms are needed to be estimated simultaneously through T daily observations during the study. The calculation is very expensive when p is large. Second, when p increases to thousands, $p(p-1)$ dramatically increases to millions, however, the daily observations within one year is only about 252. In addition, when p is very large, some of the estimation cannot be implemented correctly.

to 2015, the aggregate market value of the sample firms as a fraction of the total market value (MVS/MVT), the cross-sectional average sample firm size (FMV) and book-to-market value (BM), the proportion of each industry in the sample firms: energy (ENG), materials (MTR), industrials (IDL), consumer discretionary (CDS), consumer staples (CST), health care (HCR), financials (FNL), information technology (IT), telecommunication service (TEL), and utility (UTL), and the number of same stocks between the current and previous years (CF) are reported in the table, respectively.

[Table 2.1 inserted here]

On average, the 1,000 sample firms comprise 92% of the total market capitalization over the whole sample period from 1980 to 2015, which implies that the 1,000 sample firms are large enough to approximately represent the entire market. In fact, the aggregate market value of the 1,000 sample firms as a fraction of the total market value ranges from 89% in 1984 and during 1994–1996 to 94% during 1999–2000 and 2009–2015, which indicates that the fraction is large and stable during the sample period. Therefore, we consider the estimated Σ_t^{-1} of the 1,000 sample firms as the firm-specific network by removing the whole common information of the market, hereafter.

The cross-sectional average firm size is 7.90 billion on average over the whole sample period, increasing from 0.95 billion in 1980 to 20.37 billion in 2015. In contrast, the cross-sectional average firm book-to-market value is relatively stable over time, 0.40 on average. In addition, the proportion distribution of the industry sectors in the 1,000 sample firms varies over time. For example, the proportion of information technology increases from 6% in 1980 to 15% in 2015, while the proportion of utility decreases from 13% in 1980 to 6% in 2015.[9] Another important thing as shown in Table 2.1 is that the number of same stocks between two successive years are very large, 892 on average, with the minimum of 794 in 2000 and the maximum of 930 in 2013. This large number of same stocks ensures a sufficient number of firms for the out-of-sample applications and analyses in Section 2.5.

[9]At least 10 largest firms in each industry are included in our sample firms, which could capture large proportion of common information in the industry. More discussions will be provided in Section 3.

2.2.2 Firm-specific network

With the assumption that firm returns are multivariate Gaussian, the log-likelihood for the covariance matrix Σ_t is

$$L(\Sigma_t) \propto -\frac{pT}{2} ln(2\pi) - \frac{T}{2} ln(|\Sigma_t|) - \frac{T}{2} Trace(\Sigma_t^{-1} S_t), \tag{2.1}$$

where T is the number of daily observations within the study year t, $S_t = T^{-1} \sum_{t_i=1}^{T} (R_{t_i} - \bar{R})(R_{t_i} - \bar{R})'$, $\bar{R} = (\bar{R}_1, \cdots, \bar{R}_p)'$, and $R_{t_i} = (R_{1t_i}, \cdots, R_{pt_i})'$. As shown in (2.1), there are p^2 parameters in Σ_t needed to be estimated, i.e., the $p(p-1)$ pairwise relationships among the p firms and the variance of each stock returns in the diagonal of Σ_t. When the number of firms p is large and the number of observations T is small, i.e., $p^2 \gg T$, then Σ_t cannot be estimated by the conventional method.

In order to estimate the firm-specific network Σ_t^{-1} per year, we apply the graphical structure estimation method following [37]. The $p(p-1)$ pairwise relationships in the inverse of the covariance matrix of firm returns, Σ_t^{-1}, are estimated simultaneously, isolating the common effects between every two firms with the other sample firms in the market. Let $\Omega_t = \Sigma_t^{-1}$. Specifically, to estimate the firm-specific network Ω_t, we consider the ℓ_1-penalized log-likelihood,

$$\hat{\Omega}_t = \arg\max_{\Omega_t} \{\ln(\det(\Omega_t)) - trace(S_t \Omega_t) - \rho \|\Omega_t\|_1\}, \tag{2.2}$$

where $\|\Omega_t\|_1 := \sum_{i,j=1}^{p} |\Omega_{ijt}|$ and ρ is a fixed regularization parameter.

For each year from 1980 to 2015 inclusive, we estimate the firm-specific network Ω_t by (2.2) for the 1,000 sample firms, which allows the firm-specific relationships varying from one year to another. To implement the estimation of Ω_t in (2.2), we use the R package "huge" ([46]) and select the regularization parameter ρ in (2.2) at the first year 1980 by the method "stars" proposed in the "huge" package which gives $\rho = 0.1279$. We then use this selected ρ for the estimation of Ω_t for the following years in the sample period, which gives the same weight of penalization each year.

The estimated firm-specific network $\hat{\Omega}_t$ from (2.2) is summarized as a p×p (1,000×1,000) symmetric matrix with each element being 0 or 1. $\hat{\Omega}_{ij,t} = 0$ represents firms i and j conditionally independent given the other $p-2$ sample firms, while $\hat{\Omega}_{ij}=1$ represents firms i and j conditionally dependent. The elements in the diagonal of $\hat{\Omega}_t$ are all 1. Having demonstrated in Table 2.1, the aggregate market value of the $p=1,000$ largest sample firms accounts for more than 90% of the whole market value on average over the whole sample period from 1980 to 2015, which implies that the estimated firm-specific interrelationships are approximately the firm-specific interconnections conditional on the entire market. $\hat{\Omega}_t$, therefore, is approximate to an undirected network conditional on the entire market, describing the conditional associations among the sample firms. Each firm is taken as a node in the undirected network, with no edge ($\hat{\Omega}_{ij}=0$) implying that there is no firm-specific associations between firms i and j after the information conveyed from the entire market removed.

For each sample firm, we sum the number of interconnections in $\hat{\Omega}_t$, which describes the degree of the firm-specific connections at year t at firm-level, denoted as D_{it}, where i represents the ith sample firm, $i=1,\cdots,p$. That is,

$$D_{it}=\sum_{\substack{j=1\\ j\neq i}}^{p}\Omega_{ij,t},\quad i=1,\cdots,p \text{ and } t=1980,\cdots,2015. \tag{3a}$$

In addition, each year, we compute the total number of interconnections in $\hat{\Omega}_t$ out of the total number of all possible interconnections $p(p-1)$, which describes the aggregate degree of the firm-specific connections or the degree of the firm-specific connections at market-level, denoted as M_t. That is,

$$M_t=\frac{1}{p(p-1)}\sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p}\Omega_{ij,t},\quad t=1980,\cdots,2015. \tag{3b}$$

Note that the exact aggregate degree of the firm-specific connections is thus $M_t\times p(p-1)$ in this study.[10]

The time-series aggregate degree of the firm-specific connections (M_t) over the whole sample period from 1980 to 2015 is plotted in Figure 2.1. As can be seen from the figure, the aggregate degree of the firm-specific connections is very small, less than 4% of all possible connections

[10]The definition of M_t in (3b) facilitates the comparison of the aggregate degree of the firm-specific connections year to year if the number of sample firms varies over time.

(1000 × 999) every year over the whole sample period, without exception. In addition to the small degree, M_t exhibits an evidently downward time trend over the sample period. These imply that there is more common information than the idiosyncratic information within firms on average, and the amount of the common (idiosyncratic) information conveyed from firms is increasing (decreasing) over time.

[Figure 2.1 inserted here]

In order to examine the firm-specific connections in detail, Table 2.2 reports the aggregate degree of the firm-specific connections (M_t), the minimum, 5th percentile, 25th percentile, median, 75th percentile, 95th percentile, and maximum of the degree of the firm-specific connections D_{it} for each year from 1980 to 2015, respectively.

[Table 2.2 inserted here]

As can be seen from the table, M_t decreases from 3.43% in 1980 to 2.55% in 2015, a falling of 26% over 36 years, as illustrated in Figure 2.1. Consistent with the aggregate degree of the firm-specific connections (M_t), the individual degree of the firm-specific connections, D_{it}, also demonstrates the downward time trend. In particular, the maximum, median, and minimum of D_{it} reduce from 63, 34, and 11 in 1980 to 50, 25, and 8 in 2015, respectively. The small and declined firm-specific connections at the both aggregate-level and individual firm-level indicate that the amount of information shared among firms is large and increases over time, which is consistent with the widely believed notion that the equity market is becoming more integrated over time and firms are driven more by common factors (e.g., [2], [47]). Because M_t represents the aggregate degree of the firm-specific relationships conditional on the "market", this condition, in turn, suggests that M_t could serve as another measurement of market integration.[11] The lower the aggregate degree of the firm-specific connections (M_t) is, the more the market is integrated.

To explore the relations between the new and existing measurements of market integration, we compare the new measurement with two popular measurements: the correlation and R-square methods. First, we estimate the degree of market integration (i) by the cross-sectional average of

[11] Because the firm-specific network Ω_t is estimated using the firms within the U.S., the market integration here represents the U.S. domestic market integration, not the global market integration.

the 1000×999 pairwise correlations among the 1,000 sample firms, denoted as C_t, and (ii) by the cross-sectional average degree of firm integration measured by R-square ([2]) of the 1,000 sample firms, denoted as $R^2_{M,t}$. For each firm, to estimate the degree of its market integration, we regress the firm daily excess returns within the study year on the Fama and French five factors, the R-square from the regression measures the degree of the firm integration at that year. We find that both C_t and $R^2_{M,t}$ exhibit upward time trends, which are consistent with M_t. In particular, C_t increases from 0.188 in 1980 to 0.325 in 2015 and $R^2_{M,t}$ increases from 0.235 in 1980 to 0.395 in 2015. The correlations between C_t and M_t, $R^2_{M,t}$ and M_t are -0.661 and -0.686 over the whole sample period of 1980–2015, respectively.

We then examine whether there is any causality among these measurements of market integration. The Granger causality tests show that the decreasing in M_t leads to the increasing in C_t and $R^2_{M,t}$ at the 5% significant level, respectively.[12] However, on the other hand, the increasing in C_t or $R^2_{M,t}$ does not lead to the decreasing in M_t. In addition, there is no granger causality between C_t and $R^2_{M,t}$, but C_t and $R^2_{M,t}$ are nearly perfect correlated, with correlation 0.995 over the whole sample period. One of the possible reasons that C_t or $R^2_{M,t}$ does not granger cause M_t is that the common information embedded in C_t or $R^2_{M,t}$ is a subset of the entire market information. The detailed discussion is provided in Section 2.4. Therefore, in contrast to the correlation and R-square methods, the aggregate degree of the firm-specific connections provides a system-wide and factor-free measurement of market integration, i.e., the new measurement of integration simultaneously consider all firms in the market and does not require any risk factor to estimate R-square.

One other thing that stands out from Table 2.2 is that the aggregate degree of the firm-specific connections drops much lower during crisis years than other time. For example, M_t fell to 3.34% in 1987 (Black Monday), 3.9% and 3.26% in 1992 and 1993 (Black Wednesday), 3.13% in 1999 (Asian/Russian financial crisis), 2.77% and 2.68% in 2001 and 2002 (IT bubble burst and 9/11 incident), 2.8% and 2.74% in 2008 and 2009 (U.S. subprime mortgage crisis), 0.0281 in 2011 (European debt crisis). The substantial drops in the aggregate degree of the firm-specific connections imply

[12] The Granger causuality test is a Wald test comparing an unrestricted model with a restricted model using the lag values of the tested variable. For C_t ($R^2_{M,t}$), when one lag value of the test variable is used, the p-value of the test is 0.016 (0.012); when two lag values of the test variable are used, the p-value of the test is 0.080 (0.066). The detailed test method is discussed in Section 2.4.

that there is more common information capitalized into stock prices during crisis years, coinciding with the extant literature that firm returns tend to co-move more during crises (e.g., [48], [49]), and as a result, the degree of the firm-specific connections tends to decrease.

To formally identify the years of the substantial drop in the aggregate degree of the firm-specific connections, we apply the "Binary Segmentation" method proposed by [50] to detect the change points in the mean of M_t over the sample period. To implement the binary segmentation method, first, we multiply M_t by $1,000 \times 999$ to get the total number of conditional interconnections for the 1,000 sample firms each year from 1980 to 2015, denoted as $M_{T,t}$. We then apply the "cpt.mean" function in R package "changepoint" ([51]) to detect the change points in the mean of $M_{T,t}$ over the sample period, which are presented in Table 2.3. As demonstrated in the table, the most significant change in the mean of $M_{T,t}$ happened at 1999, followed by 1992, 2012, 2002, and 2007, which are all around crisis years.[13] This suggests that the firm-specific network is sensible in detecting the possible periods of high systemic risk and the channels of risk contagion. During the periods of high systemic risk, shocks quickly propagate within the entire financial system and more information conveyed from firms to firms, thus the number of the firm-specific interconnections falls sharply. Taken together the origins of the shocks, the firm-specific interrelationships could help tell how shocks to one firm translate into shocks to the linked firms and whether the transmission of the shocks is direct or indirect, which provides valuable information for the control and measurement of the systemic risk and the design of policy regulations.

[Table 2.3 inserted here]

On the other hand, at the firm-level, the minimum degree of the firm-specific connections D_{it} decreases during crisis years, which is consistent with M_t, whereas the maximum of D_{it} increases. For instance, the minimum of D_{it} dropped to only 5 in 1987 from 11 in 1980, but the maximum of D_{it} rose to 69 in 1987 from 63 in 1980. The opposite directions in the maximum and the minimum of D_{it} during crisis years, together with the general downward time trend of D_{it} or M_t, could

[13] It is noted that the changes in the mean of $M_{T,t}$ assume that the variance of $M_{T,t}$ does not change. As a robust check, we also examine the changes in both the mean and variance of $M_{T,t}$ by applying the "cpt.meanvar" function in the R package. Qualitatively, the results are similar to the results from the examination of changes in the mean of $M_{T,t}$ only. All the changes in the mean and variance of $M_{T,t}$ happen around crisis years. For instance, the most significant change in the mean and variance of $M_{T,t}$ happened at 1998, followed by 1992, 1982, 2012, and 2010.

suggest the least affected firms by the crisis. If firms are less affected by the market, firms have more own information, thus could have more firm-specific connections. For instance, the company "VF Corp", which is an American worldwide apparel and footwear company, has the largest degree of firm-specific connections 69 in 1987. The company "Fairchild Industries INC", which provides products in extruded rubber, molded rubber, molded plastic, extruded plastic as well as elastomer solutions for customers worldwide, has the largest degree of the firm-specific connections 60 in 1999. Firms less affected by crises could be determined by many factors, such as the nature business of the firms and the type of crises.

As a robustness check, we also examine whether the results of the firm-specific network obtained with daily observations would be altered if lower frequency observations are used instead. We apply monthly stock returns and five-year rolling windows to construct the conditional network Σ_t^{-1} per month by (2.2). Each month, beginning at January 1985, we apply the stock returns of the previous 59 months with the current month to estimate the firm-specific network for the 1,000 largest firms at the current month. The ρ in (2.2) is chosen to be the same 0.1279 as using daily observations, thus we conduct the same level of penalization as using daily stock returns. The trend of the results using monthly stock returns is in line with that using daily stock returns, i.e., the aggregate degree of firm-specific connections decreases from January 1985 to December 2015, particularly during crisis years. For example, from January 1985 to December 2015, the smallest aggregate degree of the firm-specific connections is 0.0287 at August 2013, which is estimated using the monthly stock returns from July 2008 to August 2013, covering the subprime mortgage crisis and European sovereign debt crisis. To conserve space, the results of the estimated monthly firm-specific network are not tabulated in this paper.[14]

Overall, the firm-specific network exhibits downward time trends in the degrees of the firm-specific connections at the both market-level and firm-level, which offers a new system-wide and factor-free measurement of market integration and identifies the possible periods of high systemic risk. Hereafter, throughout this paper, we measure the firm-specific network Σ_t^{-1} per year by using stock daily returns as discussed above.

[14]The detailed results are available upon request.

2.3 What Affects the Firm-specific Relationships?

Having documented the decreasing time trends in the degrees of the firm-specific connections at the both market-level and firm-level, in this section, we now turn to examine some possible explanations for the firm-specific relationships shown in Ω_t.

Let $\pi(t)$ be the symmetric probability matrix with entries $\pi_{ij}(t) = \pi_{ji}(t) = P(\Omega_{ij,t} = 1)$, for every $i = 1, \cdots, p$ and $j = 1, \cdots, p$ at year t, where $p = 1,000$ and $t = 1980, \cdots, 2015$. Therefore, $\pi_{ij,t}$ represents the probability that the ith and jth firms are conditional dependent at year t. Following [38], let $\Omega_{ij,t}|\pi_{ij}(t) \sim Bernoulli(\pi_{ij}(t))$, independently for each $i = 2, \cdots, p$ and $j = 1, \cdots, i-1$, and then we construct $\pi_{ij,t}$ by the distribution function of logistic random variables, obtaining

$$E[\Omega_{ij,t}|\pi_{ij}(t)] = \pi_{ij}(t) = \frac{1}{1 + e^{-S_{ij}(t)}}, \tag{4a}$$

and let

$$S_{ij}(t) = \mu_t + X'_{ij,t}\beta_t = \mu_t + \beta_{1,t}X_{ij1,t} + \beta_{2,t}X_{ij2,t} + \cdots + \beta_{k,t}X_{ijk,t}, \tag{4b}$$

where $X_{ij,t} = [X_{ij1,t}, \cdots, X_{ijk,t}]'$ is a k-dimensional vector of the ith and jth firms' time-varying linkages, with each element in the vector $X_{ij,t}$ describing one type of linkage between firms i and j at year t. For example, the links can be the similarities between the firm attributes, such as firm size, firm book-to-market value, industry, or headquarter location; $\beta_t = [\beta_{1,t}, \cdots, \beta_{k,t}]'$ are the corresponding dynamic effects of the linkages $X_{ij,t}$ on the firm-specific network $\Omega_{ij,t}$. It is noted from (4b) that β_t is independent of i and j, thus $\beta_{k,t}$ in β_t measures the impact of the kth type linkage among firms on the firm-specific network. Let $X^{\kappa}_{i,t} = [X_{i1\kappa,t}, \cdots, X_{ip\kappa,t}]'$, $X^{\kappa}_t = [X^{\kappa}_{1,t}, \cdots, X^{\kappa}_{p,t}]'$, $S_{i,t} = [S_{i1,t}, \cdots, S_{ip,t}]'$, $S_t = [S_{1,t}, \cdots, S_{p,t}]'$, and 1_p be a $p \times 1$ vector with one, then (4b) can be simplified into a matrix notation. That is,

$$S_t = \mu_t \times 1'_p 1_p + \sum_{\kappa=1}^{k} \beta_{\kappa,t} X^{\kappa}_t. \tag{4c}$$

Equations (4a), (4b), and (4c) describe a Bayesian nonparametric model, which can be considered as a dynamic network regression. The dependent and independent variables are all $p \times p$ matrix in the model. To estimate the effect, $\beta_{\kappa,t}$, of the κth linkage X_t^κ on the firm-specific network Ω_t, we follow the Bayesian algorithm proposed by [38], which simultaneously estimates $\beta_{\kappa,t}$ for all ts using the information from all X_t^κ, $\kappa = 1, \cdots, k$. Thus, this network regression is also a system-wide approach. The detailed prior specification and steps of posterior computation are described in the Appendix.

In the empirical study, we consider $k = 3$. X_t^1, X_t^2, and X_t^3 describe the three types of firm linkages, i.e., links through the style, industry, and headquarter location of firms. We examine whether firms from the same style, industry, or location tend to be correlated even after all the common information from the entire market removed.

To construct the $p \times p$ style matrix X_t^1 at year t, first, we identify the style membership of each sample firm. In doing so, we first classify the 1,000 sample firms into a 3×3 size and book-to-market groups using the 3×3 independent sorts on size and book-to-market values, with the breakpoints chosen as 30% and 70% percentiles of the stock size and book-to-market values, respectively. Then there are nine groups, representing the nine different styles. If firms i and j are from the same style (group) at year t, then we assign $X_{ij1,t} = 1$ in such case and $X_{ijt,t} = 0$ otherwise. Similarly, in the construction of X_t^2 and X_t^3, when firms i and j are from the same industry sector at year t, then let $X_{ij2,t} = 1$, otherwise $X_{ij2,t} = 0$; when firms i and j are headquartered in the same state at year t, then let $X_{ij3,t} = 1$, otherwise $X_{ij3,t} = 0$. Therefore, in this paper, we have,

$$S_t = \mu_t + \beta_t^S X_t^1 + \beta_t^I X_t^2 + \beta_t^L X_t^3, \tag{5}$$

where β_t^S, β_t^I, and β_t^L represent the effects of the style, industry, and headquarter location on the firm-specific interconnections at year t. The procedure in the construction of the explanatory matrixes X_t^1, X_t^2, and X_t^3 makes every X_t^κ a $p \times p$ matrix with entry 0 or 1, which facilitates the comparison among the values of β_t^S, β_t^I, and β_t^L — the effects from the style, industry, and headquarter location. The values of the estimated $\hat{\mu}_t$, $\hat{\beta}_t^S$, $\hat{\beta}_t^S$, and $\hat{\beta}_t^S$ are presented in Table 2.4 respectively, with their corresponding 95% confidence intervals.

[Table 2.4 inserted here]

As can be seen from the table, firms are more likely to be conditionally dependent when they come from the same style or industry. However, the location of firm headquarter has insignificant effect on the firm-specific connections.[15] Specifically, the industry attributes have the largest positive impact on the firm-specific interconnections, 1.197 on average over the whole sample period from 1980 to 2015, followed by the style attributes, 0.0389 on average, then the location attributes, 0.003 on average but insignificant. More specifically, the effect of industry rises from 0.808 in 1980 to 1.707 in 2015, increasing more than twice over the 36 years in the sample period. During the crisis years, $\hat{\beta}_t^I$ rose to 1.764 in 2001 and 1.725 in 2009. The larger values of the estimated $\hat{\beta}_t^I$ obtained at 2001 and 2009 thus are interpreted as indicating industry has a greater effect on the firms' interrelationships in 2001 and 2009 after all the market information removed. It is known that 2001 is the IT bubble, then the greater industry effect in 2001 might be due to the information technology sector; and 2009 is the subprime mortgage crisis, then the greater industry effect in 2009 might be due to the financials sector.

For the effect of style on the firm-specific connections, $\hat{\beta}_t^S$ rises from 0.392 in 1980 to 0.458 in 2015. Although $\hat{\beta}_t^S$ increases over the sample period and also goes up more during crisis years, the style effect increases relatively slower than the industry effect. For instance, $\hat{\beta}_t^S$ rose to 0.428 in 1987, 0.459 in 1999, 0.413 in 2008, and 0.423 in 2011. The positive and significant $\hat{\beta}_t^S$ indicates that firms belonging to the same style tend to be related with each other, which could stem from the implicit sources that determine the firm style, such as entrepreneurial skill and social network, availability of labor, and availability of finance.

In addition to the effects of style and industry, the intercept $\hat{\mu}_t$ represents the common pattern among the pairwise firm conditional interrelationships, which is negative, decreasing, and significant at least at 5% level for all 36 years over the whole sample period. Specifically, $\hat{\mu}_t$ decreases from -3.484 in 1980 to -4.029 in 2015, dropping to -4.086 in 2001 and -4.023 in 2009. The downward time trend and negative sign of $\hat{\mu}_t$ are consistent with the decreasing time trend in the aggregate

[15] The significance of β_t^I, β_t^S, and β_t^L are obtained from their corresponding confidence intervals. We also examine the 99% and 90% confidence intervals for each coefficients, besides the 95% confidence intervals reported in Table 2.4. In order to conserve space, the results are not tabulated in the paper. The detailed results are available upon request.

degree of the firm-specific connections, which provides an additional supportive evidence that firms are becoming more conditional independent over time on average. Furthermore, the magnitude of $\hat{\mu}_t$ is more than triple the value of $\hat{\beta}_t^I$ and nearly ten times the value of $\hat{\beta}_t^S$ on average over the whole sample period, which indicates the greater impact of the market on firms than firms per se, consistent with the small values of M_t.

Altogether, the results here show that the industry and style attributes positively and significantly affect the firm-specific connections to varying extents. The industry effect is always dominate the style effect and the greater effect of the industry is amplified during crises. The amplification of the industry effect during crises could relate to specific industries and the nature of crises. Beyond these attributes, many other factors could affect the firm-specific relationships, such as the products and customers of firms, the financial resources of firms, the social network of firms' managers, etc. The network regression method proposed in this study can be applied to examine the dynamic impact of these potential factors on the firms' conditional relationships when the information of these factors is available.

2.4 Implications in Asset Pricing: Known Unknown Factors

Firms' conditional relationships discussed above have immediate implications in many areas, such as asset pricing, investment, etc. For example, in asset pricing, one important issue in both theoretical and empirical validity of factor models is the correct specification of factors, including both the validity of each factor and the completeness of the factors to extract all the variations in the market. To date, a large number of different factors have been proposed.[16] In this section, we apply the firm-specific network obtained from Section 2.2 to examine the existence of unknown common factors given a set of well-known risk factors. Specially, for the examination, we consider the set of the Fama-French five factors ([7]), and then add the momentum factor ([8]) to the set.

2.4.1 Existence of unknown factors

With the assumption that firm returns are multivariate Gaussian, the partial correlation between two firms is zero if and only if they are conditionally independent given the entire market ([39]).

[16] See, for example, [17], [8], [52], [7], and [53].

Therefore, if a set of existing common factors captures all common information in the market, then the cross-sectional average partial correlation among firms should display a downward time trend same as the aggregate degree of the firm-specific connections, with the effect of the set of factors on the firms' relationships removed. If, in contrast, the cross-sectional average partial correlation among firms does not exhibit the same downward time trend as the aggregate degree of firm-specific network, then the differential in the trends implies one or more common factors omitted from the given set of factors.

Let $\varrho_{ij,t}$ represent the partial correlation between the ith and the jth firms at year t given a set of known factors. In particular, to compute $\varrho_{ij,t}$ and then examine the possible existence of unknown factors, we first regress the individual stock daily excess returns on the five factors of Fama and French (FF, hereafter): market factor (MRP), size factor (SMB), value factor (HML), profit factor (RMW), and investment factor (CMA), and get the residuals for each sample firm. The FF five factors and the U.S. risk-free rate are obtained from French's website. We then compute the sample partial correlation by calculating the correlations between the residuals from the regressions for firms i and j.

Beginning from 1980, the excess returns of each individual stock in our sample are regressed on the FF five factors per year over the whole sample period,

$$\begin{aligned} R_{it,d} = \alpha_t &+ \beta_t^{MKT} MKT_{t,d} + \beta_t^{SMB} SMB_{t,d} \\ &+ \beta_t^{HML} HML_{t,d} + \beta_t^{RMW} RMW_{t,d} + \beta_t^{CMA} CMA_{t,d} + \epsilon_{it,d}, \end{aligned} \tag{6}$$

where $R_{it,d}$ is excess return of the $i$the firm at year t and day d, $d = 1, \cdots, T$, and T is the total number of days in year t. Thus, $\varrho_{ij,t} = corr(\epsilon_{it}, \epsilon_{jt})$, where $\epsilon_{it} = [\epsilon_{it,1}, \cdots, \epsilon_{it,T}]'$. The cross-sectional average $\varrho_{ij,t}$, denoted as $\bar{\varrho}_t$, describes the aggregate firm partial correlation given the five factors of FF. That is,

$$\bar{\varrho}_t = \frac{1}{p(p-1)} \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \varrho_{ij,t}, \quad p = 1,000 \text{ and } t = 1980, \cdots, 2015.$$

The time series values of $\bar{\varrho}_t$ are reported in the 10th columns of Table 2.2 and are plotted in Figure

2.1, in contrast with the aggregate degree of the firm-specific connections.

As shown in Table 2.2, the aggregate firm partial correlation ($\bar{\varrho}_t$) given the FF five factors increases over time, particularly during crisis years, which exhibits an upward time trend opposite to M_t as illustrated in Figure 2.1. Specifically, $\bar{\varrho}_t$ rises from 0.0582 in 1980 to 0.0739 in 2015. During crisis years, for instance, $\bar{\varrho}_t$ rose to 0.0688 in 1987, 0.0712 in 2001, and 0.0826 in 2008. In addition, the greatest significant change year in the mean of $\bar{\varrho}_t$ is 1999 over the whole sample period of 1980–2015, which is same as M_t, followed by 2006, 2009, 2002, and 1997. Similar to the change points in the mean of M_t, all the significant changes in the mean of $\bar{\varrho}_t$ over the whole sample period happen around crisis years, however, $\bar{\varrho}_t$ is soaring instead of decreasing as M_t. But due to the "if and only if" relationship between $\varrho_{ij,t}$ and $\Omega_{ij,t}$, if the five factors of FF capture all the common information of the entire market, $\bar{\varrho}_t$ would display the same downward time trend as M_t and fall sharply during crisis years. Therefore, the opposite directions of the time trends between $\bar{\varrho}_t$ and M_t, in turn, indicate that the FF five factors cannot represent all the risk factors, there should exist some other common risk factors.

Another intriguing thing is: the cross-sectional average variance of the residuals from the regressions in equation (6), denoted $\sigma^2(\epsilon_{it})$ as the residual variance of the ith firm, is generally decreasing or stable during the whole sample period of 1980–2015 but increases sharply during crises. For example, the time-series mean of the cross-sectional average $\sigma^2(\epsilon_{it})$ is 4.285 over the whole sample period, but the cross-sectional average $\sigma^2(\epsilon_{it})$ was 16.168 in 2000 and 8.909 in 2008. Thus, during crisis years, both the variance and the pairwise correlation among the residuals go up substantially, which result in a quite large increase in the covariance of the residuals. The rising covariance of the residuals then implies the co-movements of the residuals and further indicates the possible existence of unknown common factors besides the FF five factors.

The FF five factors do not contain the momentum factor, which is another well-known common risk factor and frequently used in asset pricing models. To further investigate the existence of unknown common factors, we add the momentum factor to the FF five factors in (6) and examine whether the aggregate partial correlation would display a decreasing time trend,

$$\begin{aligned} R_{it,d} = \alpha_t &+ \beta_t^{MKT} MKT_{t,d} + \beta_t^{SMB} SMB_{t,d} \\ &+ \beta_t^{HML} HML_{t,d} + \beta_t^{RMW} RMW_{t,d} + \beta_t^{CMA} CMA_{t,d} + \beta_t^{MOM} MOM_{t,d} + \epsilon_{it,d}. \end{aligned} \tag{7}$$

The cross-sectional average partial correlation among the 1,000 sample firms estimated using the new six-factor regression again demonstrates an upward time trend over the whole sample period. In particular, $\bar{\varrho}_t$ estimated from (7) increases from 0.0557 in 1980 to 0.0662 in 2015. During crisis years, it rises sharply, similar to $\bar{\varrho}_t$ estimated from (6) given the FF five factors. For example, given the set of the six factors in equation (7), $\bar{\varrho}_t$ rose to 0.0685 in 1987, 0.0703 in 2002, and 0.0799 in 2008. Thus, the results from using the six factors are consistent with the results found by using the FF five factors and indicate that there is at least one common risk factor omitted from the set of the FF five factors and the momentum factor.

2.4.2 Test for the existence of unknown factors

In order to formally test the existence of unknown factors beyond the FF five factors, we start with the "if and only of" relationship between M_t and $\bar{\varrho}_t$. [39] shows that with the assumption that firm returns are multivariate Gaussian, the relationship between the partial correlation $\varrho_{ij,t}$ ($\bar{\varrho}_t$) and the conditional independence $\Omega_{ij,t}$ (M_t) is "if and only if", i.e. $M_t \Leftrightarrow \bar{\varrho}_t$ for any t.

We then employ the Wald test (Granger causality test) by comparing the unrestricted model in which $\bar{\varrho}_t$ is explained by the lags (up to order L) of $\bar{\varrho}_t$ and M_t, and the restricted model in which $\bar{\varrho}_t$ is only explained by the lags (up to order L) of $\bar{\varrho}_t$, that is

$$\bar{\varrho}_t = \sum_{l=1}^{L} \bar{\varrho}_{t-l} + \sum_{l=1}^{L} M_{t-L}, \qquad \bar{\varrho}_t = \sum_{l=1}^{L} \bar{\varrho}_{t-l} \tag{9a}$$

If $\bar{\varrho}_t \Rightarrow M_t$ for any t, an insignificant result is expected from (9a). Similar to the unrestricted and restricted models in (9a), we also compare the following unrestricted and restricted models,

$$M_t = \sum_{l=1}^{L} M_{t-l} + \sum_{l=1}^{L} \bar{\varrho}_{t-L}, \qquad M_t = \sum_{l=1}^{L} M_{t-l} \tag{9b}$$

If $M_t \nRightarrow \bar{\varrho}_t$ for any t, an insignificant result is expected from (9b).

For the comparison of models in (9a), the F-statistics from the Wald test shows $\bar{\varrho}_t \nRightarrow M_t$, instead, M_t granger causes $\bar{\varrho}_t$ statistically significant at least at 10% level. In particular, when $L = 1$, the F-statistics is 5.108 with the p-value 0.031. When $L = 2$, the F-statistics is 3.027 with the p-value 0.064. One the other hand, when comparing models in (9b), the F-statistics from the Wald test shows the insignificance at all lags, indicating $M_t \nRightarrow \bar{\varrho}_t$. This again supports our prior finding that there exist some unknown factors, at least one, besides the FF five factors, otherwise the directions tested in both (9a) and (9b) should be statistically insignificant. The results here indicate that the information extracted by the five factors of FF is a subset of the information embedded in M_t, thus M_t granger causes $\bar{\varrho}_t$. We also apply the same method to $\bar{\varrho}_t$ obtained from equation (7) and the same conclusion is obtained.

Overall, the comparison between the firm-specific network and the partial correlations provides a diagnostic criterion for the adequacy of existing factors in a given factor model. The criterion does not depend on a factor structure of the given factor model — no matter the factor structure is time-invariant or time-varying, and determines whether there is at least one common factor omitted in the given set of factors to extract all systematic risk, which assists in improving the theory of asset pricing. In addition, it is noteworthy that the diagnostic criterion proposed in this study is applicable not only to financial factor models, but also to other models, such as macroeconomic factor models.

Then what could be the possible unknown common factors? A good number of papers propose different risk factors. For example, there are a lot of studies on the idiosyncratic risk in the previous literature. [54] find a noticeable upward trend in the idiosyncratic volatility relative to the market volatility, which helps to predict the GDP growth. [55] show that firms' idiosyncratic volatility obeys a strong factor structure and shocks to the common idiosyncratic volatility factor are priced. These indicate that the idiosyncratic factor might exist. In addition, in the FF five-factor model ([7]), profit and investment factors might capture part of the industry information, but no factor extracts the whole industry information. The prior findings in Section 2.3 show that firms from the same industry tend to relate with each other even after all common information removed. [56] finds that stocks in economically related supplier and customer industries cross predict each other's returns. These indicate that the industry factor might exist. Although many different risk factors are proposed and studied, it seems there is no conclusive full set of factors. The firm-specific network proposed in this study helps assess the adequacy of a set of factors in extracting all variations in the market. Many issues in factor models of asset pricing await for further research which are beyond the scope of this paper.

2.5 Implication in Risk Diversification

As shown in the preceding sections above, a large number of firms are conditional independent given all the other sample firms per year and the number of these conditional independent firms rises over time. If firms are conditional independent, then the partial correlation between the firms is zero. Therefore, if investors can consistently identify the conditional independent stocks and diversify with these stocks, they may benefit consistently with few stocks in terms of risk reduction and low transaction cost.

To test this conjecture, we first cluster all the sample firms into different communities per year based on the firm-specific network Ω_t, which describes the conditional independence/dependence structures among all the sample firms. Thus, firms within the same community are more conditionally linked by firm-specific properties than firms across different communities. We then compare the effectiveness of risk diversification with three groups of stocks, i.e., stocks only within a community, stocks only across communities, and stocks both within and across communities, and examine the performance of the portfolios consisted of the stocks from the most effective risk diversification group.

2.5.1 Firm clustering

To uncover the community structure in the estimated firm-specific network of the 1,000 sample firms, we apply the fast greedy community detection algorithm proposed by [43]. The algorithm is agglomerative. At each step, two groups merge and the merging is decided by optimising the modularity, which is used to measure the strength of communities. Network with a high modularity has dense connections between the nodes within communities but sparse connections between nodes in different communities. Thus, a high modularity is interpreted as indicating tight connections among firms within the same community, but loose connections among firms across different communities.

To implement the fast greedy algorithm, we apply the R package "igraph"([57]).

Each year over the sample period from 1980 to 2015, we cluster the 1,000 sample firms into different communities based on $\hat{\Omega}_t$. The time-series modularity of the communities is plotted in Figure 2.2. As can be seen from the figure, the modularity exhibits an upward time-trend over the sample period of 1980-2015, especially during crisis years. This implies that the firm-specific connections become tight within the same community over time and especially during crises, although the number of the firm-specific connections decreases over time and drops sharply during crises.

[Figure 2.2 inserted here]

To examine the community structure among firms in detail, we study the number of communities, the modularity of the communities, and the number of firms within each community per year over the sample period of 1980–2015. The results are reported in Table 2.5.

[Table 2.5 inserted here]

As can be seen from the table, the number of communities for the 1,000 sample firms based on the firm-specific network is small, either four or five for 32 years out of the 36 years in the sample period, with six communities grouped for the other four years. This implies that firms can be stably clustered into four or five groups based on the firm-specific relationships. This small number of communities may help study the idiosyncratic risk. Firms within a community have more firm-specific properties in common than firms across communities, then understanding which properties in common within each community may assist in constructing a factor capturing the idiosyncratic risk.

In addition, Table 2.5 shows that the modularity of the communities increases more than 40% over the sample period, from 0.218 in 1980 to 0.306 in 2015, and went up to 0.262 in 1987, 0.359 in 2001, and 0.307 in 2008. The rising modularity indicates the increasing associations among firms within the same community, particularly during crisis years. Moreover, for the majority of the 36 years in the sample period, only three communities from all the communities have a large number of firms each year. For example, there are four communities detected in 1981 with 372, 344, 261, and 23 firms within each community, respectively; there are five communities detected in 1983 with 397, 342, 232, 18, and 11 firms within each community, respectively. The community structure based on the firm-specific network, then, could be employed to identify the channels of risk transition, i.e., the direct and indirect risk propagation – within and cross communities, and detect the center and bridge of contagion – link firms between communities, which can provide useful information for policy designs and regulations.

2.5.2 Comparison of risk diversification

Next, we study the effectiveness of conditional independent stocks in risk diversification in this subsection. Following [44], we use the simulation method to examine the potential gain in terms of risk diversification by three groups of stocks: stocks only within the same community, stocks only across different communities, and stocks both within and across communities.

To implement the simulation method for the examination, each year, from 1980 to 2015, we first randomly choose a varying number of stocks (ranging from 1 to 20) with replacement from each community. We then compute the risk deduction by the three groups of stocks per year. For the risk deduction by stocks only within the same community, we first apply the selected stocks from each community to form an equal-weighted portfolio and compute the variance of the monthly portfolio returns over the year for each community. We then take an average of the portfolio variances for all the communities, denoted as $\sigma^2_{IC,i,t}$, which represents the risk deduction by the stocks within the same community only at year t when i number stocks are selected from each community. For the risk deduction both within and across communities, we first apply the selected stocks from all the communities to form one equal-weighted portfolio (IP) and then compute the variance of the monthly portfolio returns over the year, denoted as $\sigma^2_{IP,i,t}$, which represents the risk deduction by the stocks both within and across communities at year t when i number stocks selected from each community. When $i = 1$, $\sigma^2_{IP,1,t}$ represents the risk deduction using the stocks across communities only. As shown in Table 2.5, the number of communities is at most six, when $i = 1$, the equal-weighted portfolio IP is constructed using at most six stocks. Thus, for simplicity, we only consider selecting stocks across all the communities when examining the risk deduction by stocks only across different communities.

The subscript "I" in $\sigma^2_{IC,i,t}$ and $\sigma^2_{IP,i,t}$ represents that the analysis is in-sample, and the subscript "C" and "P" in $\sigma^2_{IC,i,t}$ and $\sigma^2_{IP,i,t}$ represent within the same community only, and within and across communities, respectively. Beside the in-sample analysis, we also examine the out-of-sample risk deduction by the three groups of stocks. Each year, from 1980 to 2014, we apply the same simulation method described above, but compute the variance of monthly portfolio returns over the next year as the out-of-sample portfolio variance. Let $\sigma^2_{OC,i,t}$ and $\sigma^2_{OP,i,t}$ denote the risk deduction by stocks within the same community only and by stocks both within and across communities at year $t+1$, respectively. The subscript "O" in $\sigma^2_{OC,i,t}$ and $\sigma^2_{OP,i,t}$ represents that the analysis is out-of-sample.

We repeat the above procedure 5,000 times for each year and compute the simulation average of $\sigma^2_{IC,i,t}$, $\sigma^2_{IP,i,t}$, $\sigma^2_{OC,i,t}$, and $\sigma^2_{OP,i,t}$, respectively, for each i, $i = 1, \cdots, 20$. Without confusion, the simulation average for each type of risk deduction is denoted the same as $\sigma^2_{IC,i,t}$, $\sigma^2_{IP,i,t}$, $\sigma^2_{OC,i,t}$, and $\sigma^2_{OP,i,t}$, respectively. Figure 2.3 plots the time-series average of $\sigma^2_{IC,i,t}$, $\sigma^2_{IP,i,t}$, $\sigma^2_{OC,i,t}$, and $\sigma^2_{OP,i,t}$ for $i = 1, \cdots, 20$. Figure 2.4 plots the time-series $\sigma^2_{IC,20,t}$ and $\sigma^2_{IP,20,t}$, illustrating how the portfolio risk behaves over time.[17]

[Figure 2.3 inserted here]

As can be expected, using stocks across different communities are much effective in reducing the portfolio risk than using stocks within communities, as illustrated from Figure 2.3. Specifically, when selecting only one stock from each community, the time-series average of $\sigma^2_{IP,1,t}$ is 0.0091 which is close to the time-series average of $\sigma^2_{IC,20,t}$, 0.0090, selecting 20 stocks from each community. This indicates that fewer stocks from different communities are needed to reduce the risk to a given level of diversification than from the same community. Furthermore, $\sigma^2_{IP,i,t}$ and $\sigma^2_{OP,i,t}$ are much smaller than $\sigma^2_{IC,i,t}$ and $\sigma^2_{OC,i,t}$ for all is, respectively, without exception. $\sigma^2_{IP,i,t}$, $\sigma^2_{OP,i,t}$, $\sigma^2_{IC,i,t}$, and $\sigma^2_{OC,i,t}$ all converge when then number of stocks selected per community increases up to around 12. However, the convergence level of $\sigma^2_{IC,i,t}$ ($\sigma^2_{OC,i,t}$) is much higher than the level of $\sigma^2_{IP,i,t}$ ($\sigma^2_{OP,i,t}$). This again implies that holding stocks across different communities does effectively diversify risk on average over time.

[17]In order to conserve space, the results are not tabulated in the paper. The detailed results are available upon request.

To further examine the dynamic behavior of risk diversification by using stocks across different communities, we study the time-series behavior of $\sigma^2_{IC,20,t}$ and $\sigma^2_{IP,20,t}$ for the 36 years over the sample period 1980–2015. When 20 stocks are randomly chosen from each community with replacement, we assume the idiosyncratic risk can be diversified adequately. As can be seen from Figure 2.4, the portfolio risk $\sigma^2_{IC,20,t}$ is always larger than $\sigma^2_{IP,20,t}$ for all years without exception. $\sigma^2_{IC,20,t}$ increases much higher during crisis years, however, by holding stocks across different communities, the portfolio risk $\sigma^2_{IP,20,t}$ is reduced noticeably during the crisis years, especially 2001. Taken together, the results of risk diversification based on the clustered communities from the firm-specific network suggest that investors could benefit significantly from "selective" stocks focusing on both the conditional dependent (within communities) and the conditional independent (across communities) relationships.[18]

2.5.3 Portfolio performance

Having documented the benefit of risk diversification using the conditional dependent and independent stocks, we now turn to examine the performance of the resulting equal-weighted portfolios, both in-sample and out-of-sample cases, constructed from stocks both within and across communities.

To investigate the performance of portfolios, for each replication among all the 5,000 replications, we first compute the Sharpe ratio (SHP, hereafter) for each portfolio over the whole sample period, which is 1980–2015 for the in-sample case and 1981–2015 for the out-of-sample case, and then compare the portfolio SHP with the SHP of the value-weighted market index. The monthly returns of the value-weighted market index are obtained from CRSP. Panel A of Table 2.6 summarizes the simulation average of the time-series means and standard deviations of portfolio returns, the simulation average and standard deviation of portfolio SHPs, the proportion that the portfolio SHP is higher than the index SHP among the 5,000 replications, and the lower and upper bounds of the 95% confidence interval for the portfolio SHPs for both in-sample and out-of-sample cases, respectively.

[18]Because the number of communities is small, the level of risk diversification by stocks "only" across different communities is restricted.

[Table 2.6 inserted here]

As shown in Panel A of Table 2.6, the resulting equal-weighted portfolios, constructed from randomly selecting 1 to 20 stocks per community, exhibit remarkably stable portfolio returns but monotonically decreasing standard deviations, and can consistently yield higher SHPs than the market index. Specifically, for the in-sample analysis, the simulation average of the time-series mean of the resulting portfolio returns over the sample period of 1980–2015 ranges from 1.337 percent when four stocks chosen per community to 1.341 percent when two stocks chosen per community, with 1.339 percent on average, whereas the simulation average of portfolio returns' standard deviations decreases from 6.760 when one stock randomly selected per community to 5.239 when 20 stocks randomly selected per community.

When only one stock randomly selected from each community, the resulting equal-weighted portfolio yields a monthly SHP of 0.143 on average over the 5,000 replications from 1980 to 2015 and has larger SHP than the market index 58.9% times out of the 5,000 repetition.[19] When the number of stocks randomly selected from each community with replacement increases to five, which implies at most 20 to 30 stocks selected in total per year, the resulting equal-weighted portfolio yields a monthly SHP of 0.175 on average and outperforms the market index 98.8% times out of the 5,000 replications in terms of higher SHP, close to 100%. Moreover, as indicated from the 95% confidence interval of the portfolio SHPs, when at least five stocks randomly selected per community with replacement, the simulation average of the resulting portfolios' SHPs is significantly higher than the SHP of the market index at least at 5% significance level. As the number of stocks selected per community increases from 1 to 20, the range of the 95% confidence interval of the portfolio SHPs decreases more than 70%, from 0.122 to 0.036, which indicates the increasing significance of the portfolio performance (SHP) compared with the market index. When the number of stocks randomly selected from each community with replacement increases to ten, the resulting equal-weighted portfolio can 100% times beat the market index among all the 5,000 replications in terms of larger SHP, without exception.

For the out-of-sample analysis, the risk deduction from the within and across communities

[19] The monthly Sharpe ratio of the value-weighted market index is 0.136 over the period 1980–2015. The monthly Sharpe ratio of SP500 index is 0.093 over the period 1980–2015 and 0.089 over the period 1981–2015, which is lower than the value-weighted market index.

selection is also very effective. For instance, when four stocks randomly chosen from each community with replacement, which means at most 16 to 24 stocks selected in total, the resulting equal-weighted portfolio has a monthly SHP of 0.131 on average and yields larger SHP than the market index more than 50% times out of the 5,000 replications over the period 1981 – 2015. If we increase the number of stocks selected from each community with replacement to 19, the resulting equal-weighted portfolio has a monthly SHP of 0.139 on average and yields larger SHP than the market index more than 80% times out of the 5,000 replications. It is noteworthy that we simply use an equal-weighted portfolio and a randomly selection, not an optimal portfolio. Thus, the performance of the resulting equal-weighted portfolio is relatively conservative and might be considerably improved by finding the optimal weights for each stock.

To further examine the dynamic behavior of the resulting equal-weighted portfolio performance by holding stocks within and across communities and also to ensure the robustness of the performance of the resulting portfolios, we evenly divide the whole sample period into two sub-sample periods, 1980–1997 and 1998–2015, and then repeat the analysis of the resulting portfolio performance during the two sub-sample periods separately. The summaries of the results from two sub-sample periods' studies are reported in Panel B and Panel C of Table 2.6, respectively. Qualitatively, we obtain similar results to the results reported in Panel A of Table 2.6. During the early sub-sample period, for instance, when nine stocks are randomly selected per community per year, the resulting equal-weighted portfolio yields a monthly SHP of 0.204 on average and wins the market index 90.3% times out of the 5,000 replications for the in-sample case, and yields a monthly SHP of 0.149 but betas the market index only 11.34% times out of all the replications for the out-of-sample case. During the later sub-sample period, for instance, when seven stocks are randomly selected per community per year, the resulting equal-weighted portfolio yields a monthly SHP of 0.177 and beats the market index 99.98% times out of all the 5,000 replications for the in-sample case, and yields a monthly SHP of 0.114 and betas the market index 90.46% times out of all the replications. The results from the two sub-sample periods analysis indicate that the potential gain by holding conditional independent stocks is stronger and more consistent over the later sub-sample period than the earlier sub-sample period. This is likely to be due to the increased correlations among firms, reflecting the market integration. The rising correlations among firms, then, largely eliminate room for gainful diversification from the conventional method — holding a portfolio of a large number of stocks, and the number of stocks needed to achieve a given level of diversification has to be increased by applying this conventional method.

In sum, the results from the analysis of portfolios' performance indicate that investors can consistently gain from holding few stocks both within and across communities.[20] The gain is robustness to the choice of stocks and time, i.e., random selection within and across communities, and ex-post and ex-ante. Despite the correlations among firms go up over time, the conditional independent structure embedded in the firm-specific network allows investors consistently gain from risk diversification with few stocks. In addition, this potential consistent gain becomes strong over time.

2.5.4 Conditional centrality versus noncentrality firms

Besides the conditional independent structure embedded in the firm-specific network, the degree of the firm-specific connections indicates the conditional centrality structure of firms. That is, firms conditionally connected with more other firms are conditionally more central than the other firms. These central firms might bear more firm-specific risk, not only the risk per se but also the risk from the conditionally connected firms. Moreover, firms with large degrees of the firm-specific connections might come from some specific industries. In this subsection, we then examine whether the degree of the firm-specific connections relates to the industry classification and has impact on the stock performance.

Each year, from 1980 to 2015, we first sort the 1,000 sample stocks by D_{it} and form quintile equal-weighted portfolios over that year. Portfolio 1 has sparse connections, i.e., the fewest degree of the firm-specific connections, while portfolio 5 has dense connections, i.e., the largest degree of the firm-specific connections. In addition to these in-sample equal-weighted portfolios, we also construct the out-of-sample portfolios using the stocks monthly returns over the next year.

For each industry sector, the percentage of firms from the 1,000 sample firms in each portfolio

[20] We also examine the αs from regressions the resulting equal-weighted portfolio excess returns on FF five factors. We find that most of the αs from the 5,000 replications are positive for the in-sample analysis and majority of these αs are significant at least at 10% level when at least 16(8) stocks randomly selected from each community with replacement over the whole (later sub) sample period. However, for the out-of-sample analysis, most of the αs become negative but insignificant over the earlier sub-sample period, positive but insignificant over the later sub-sample. The detailed results are available upon request.

belonging to the industry sector is reported in Table 2.7. The time-series average and standard deviation of each portfolio return and SHP are reported in Table 2.7 for both in-sample and out-of-sample analysis, respectively.

[Table 2.7 inserted here]

As can be seen from Panel A of Table 2.7, firm industry membership is related to the degree of the firm-specific interconnections, which is consistent with the prior findings in Section 2.3. For example, 37.45% of energy firms, 28.86% of consumer staples firms, and 28.05% of telecommunication services are in portfolio 1, indicating that firms belonging to energy, consumer staples, telecommunication services are less likely to be conditional interrelated with each other. On the other hand, 31.29% of industrials firms, 26.80% of information technology firms, and 25.75% of financials firms are in portfolio 5, indicating that firms belonging to industrials, information technology, and financials are more likely to be conditional interrelated with each other.

As for the stock performance shown in Panel B of Table 2.7, the firms with large (small) number conditional connections tend to have high (low) return and high (low) risk. For the in-sample centrality sorted portfolios, both the time-series average return and standard deviation increase monotonically from portfolio 1 to portfolio 5. For instance, the time-series average monthly returns of portfolio 1 and 5 are 1.2% and 1.7% with standard deviation 0.043 and 0.060, respectively, which result in the monthly SHPs of 0.183 and 0.215. For the out-of-sample centrality sorted portfolios, the time-series average returns of portfolios are close to each other, but the time-series average standard deviations increase from portfolio 1 to portfolio 5, same as in-sample results. For example, the time-series average returns of portfolio 1 and 5 are 1.0% and 1.1% with standard deviation 0.048 and 0.058, respectively, resulting in the monthly SHP of 0.136 and 0.123. The higher return and standard deviation of the conditionally central firms imply that these central firms bear not only the risk per se, but also the specific risk from the connected firms, besides the common (systematic) risk from the market.

2.6 Conclusion

Firms do not exist as independent entities, but are complexly linked to each other through many direct and indirect, implicit and explicit relationships. These relationships together result in the network of firms, which is affected by firms per se and by the market. In this study, we propose a system-wide approach to estimate the firm-specific network by removing the effect from the market. The time trend in the estimated firm-specific network, on the other hand, indicates the increasing positive effect of the market on the firms' relationships, which is consistent with the well-known market integration.

In addition, the estimated firm-specific network describes the firm-specific relationships which has important implications in the study of market integration, systemic risk, asset pricing, and risk diversification. The aggregate degree of the firm-specific network provides a system-wide measurement of market integration, without dependence on any common risk factors. The time trend in the aggregate degree of the firm-specific network indicates the possible periods of the high systemic risk. The comparison between the downward time trend in the aggregate degree of the firm-specific network and the upward time trend in the cross-sectional average partial correlation among firms given a set of known risk factors implies the existence of unknown common risk factors, which assist in improving the theory of asset pricing. However, determining which are those unknown factors is a complex issue and is beyond the scope of this paper. Further research are needed on this problem. Lastly, investors can consistently benefit from the risk diversification by holding few conditional independent stocks. The consistence in this gainful diversification is not only consistent over time, but also consistent on the choices of assets. In other words, investors can benefit both ex ante and ex post, and through randomly selecting conditional independent stocks.

Beyond all of these, there are many more implications of the firm-specific network. For example, the examination of the conditional network across different markets (e.g., stock market, bond market, commodity market, currency market, or financial derivative market), the construction of the idiosyncratic factor through the communities clustered from the firm-specific network, the diagnosis of omitted factors in macroeconomic models, etc. We hope that the method proposed in this study sheds light on the study of the conditional relationships and their implications in different areas of finance.

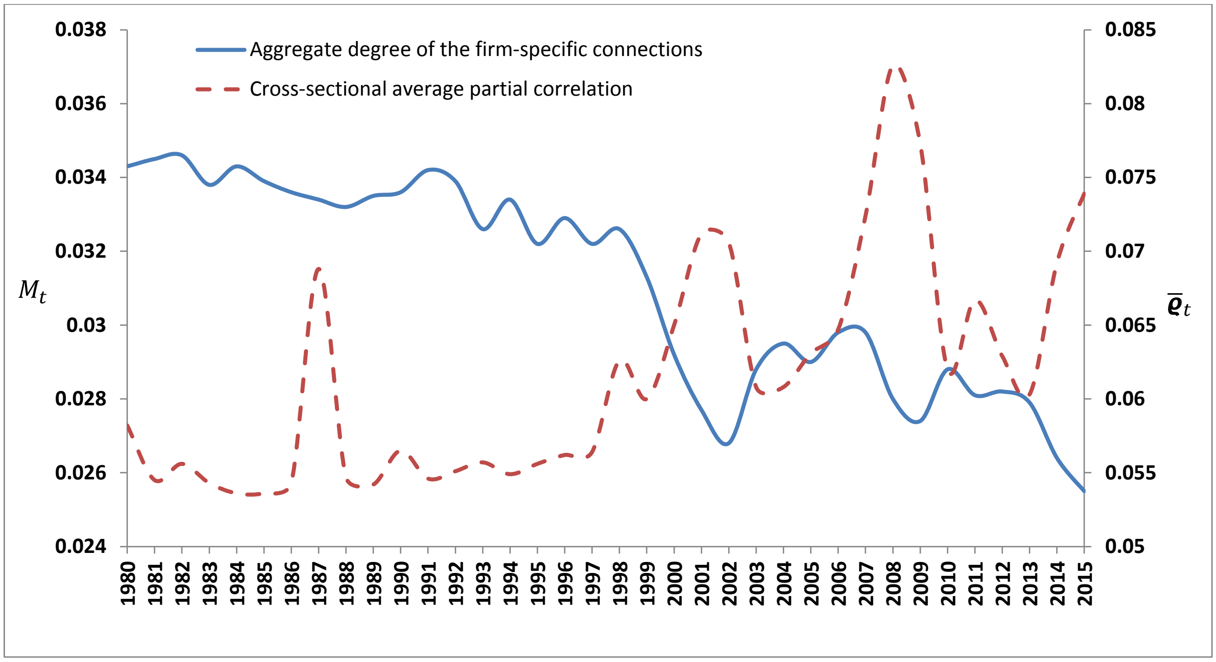

Figure 2.1: Aggregate firm-specific connections versus partial correlations.
The figure plots the aggregate degree of the firm-specific network M_t and the cross-sectional average of the partial correlations ϱ_t given Fama and French five factors among the 1,000 sample firms over the sample period from 1980 to 2015, respectively. The solid line represents the aggregate degree of firm-specific network, corresponding to the left y-axis. The dashed line represents the cross-sectional average of the partial correlations, corresponding to the right y-axis.

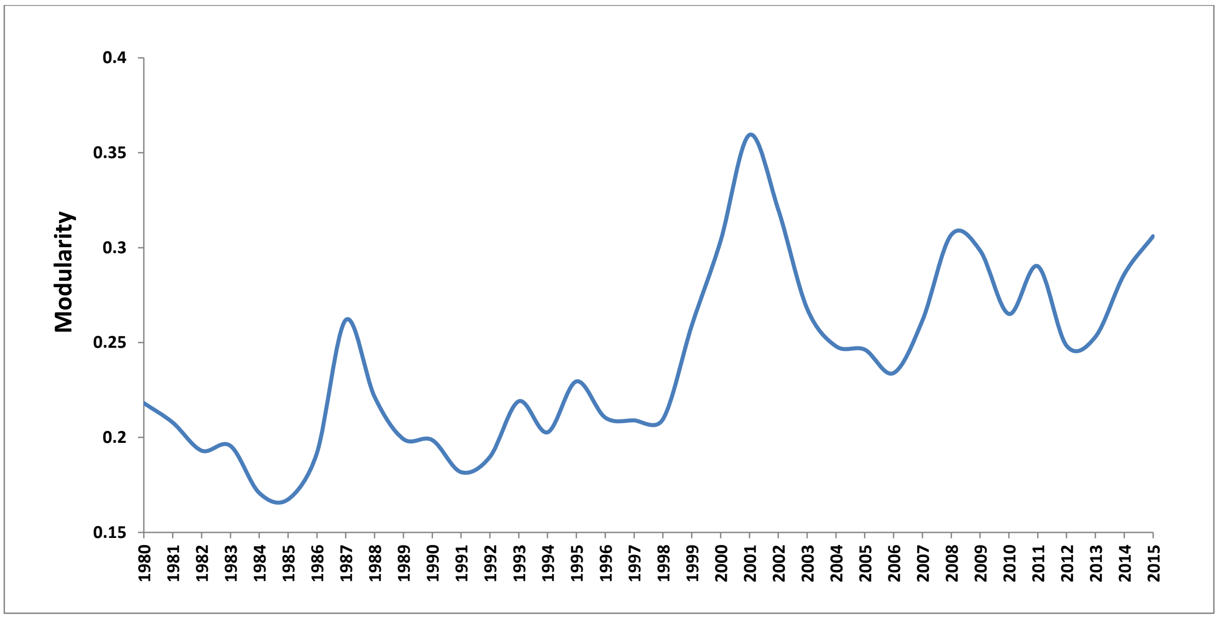

Figure 2.2: Modularity of the communities clustered from the firm-specific network. This figure plots the modularity of the communities clustered from the firm-specific network per year over the sample period from 1980 to 2015, which measures the tightness of the communities. A high modularity indicates tight connections among firms within the same community, but loose connections among firms across different communities.

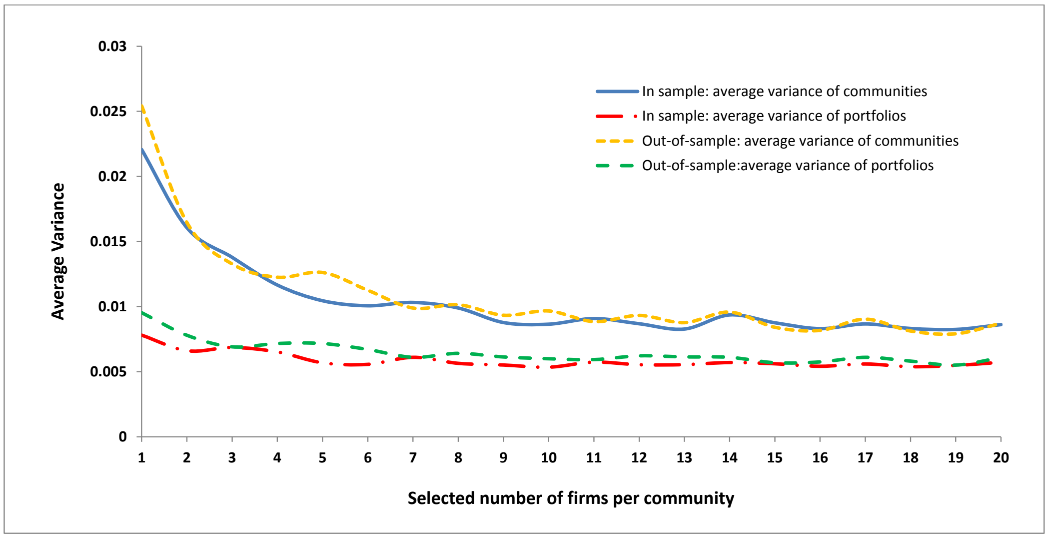

Figure 2.3: Average community variance and portfolio variance.
This figure plots the average of the time-series mean of the cross community average variance ($\sigma^2_{IC,i,t}$, $\sigma^2_{OC,i,t}$) and the average of the time-series mean of the portfolio variance ($\sigma^2_{IP,i,t}$, $\sigma^2_{OP,i,t}$) over the 5,000 simulation replications, for both in-sample and out-of-sample studies, respectively. The solid lines represent the results of in-sample study and the dashed lines represent the results of out-of-sample study.

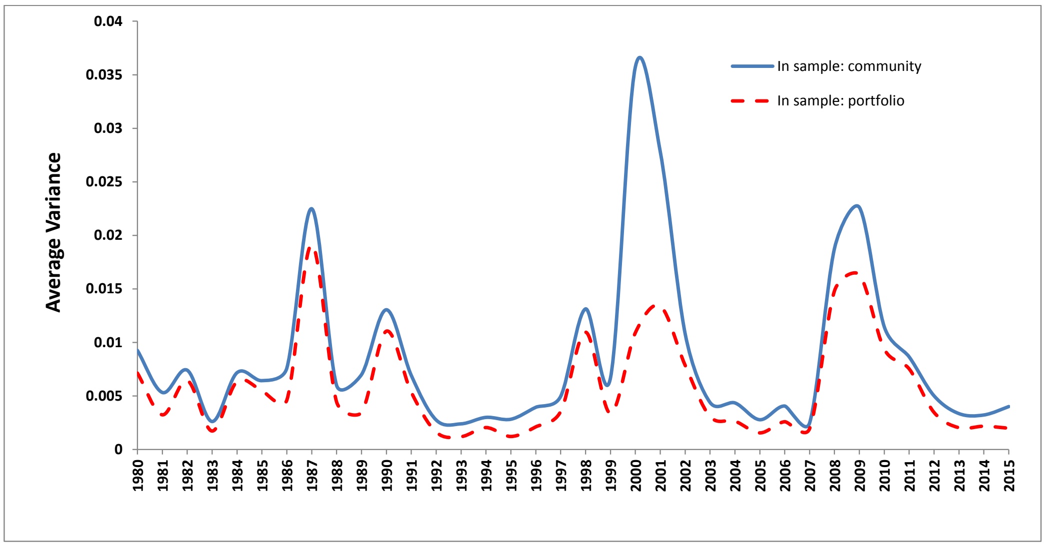

Figure 2.4: Time-series community variance and portfolio variance.
This figure plots the average of the in-sample time-series cross community average variance ($\sigma^2_{IC,20,t}$) and the portfolio variance ($\sigma^2_{IP,20,t}$) for 20 firms randomly chosen from each community with replacement over the 5,000 simulation replications, respectively. The solid line represents $\sigma^2_{IC,20,t}$ and the dashed line represents $\sigma^2_{IP,20,t}$.

Table 2.1: Summary statistics.

Each year, from 1980 to 2015, the aggregate market value of the 1,000 sample firms as a fraction of the total market value (MVS/MVT), the cross-sectional average sample firm size (FMV) and book-to-market value (BM), the proportion of each industry sector in the 1,000 sample firms: energy (ENG), materials (MTR), industrials (IDL), consumer discretionary (CDS), consumer staples (CST), health care (HCR), financials (FNL), information technology (IT), telecommunication service (TEL), and utility (UTL), and the number of same stocks between the current and previous years (CF), are reported in the table, respectively. Firm size is computed by firm's market value of equity. Firm book-to-market value is computed by firm's common equity over total asset.

Year	MVS/ MVT	FMV ($ B)	BM	ENG	MTR	IDL	CDS	CST	HCR	FNL	IT	TEL	UTL	CF
1980	0.92	0.95	0.42	10	11	20	14	9	5	10	6	2	13	
1981	0.91	1.03	0.43	10	11	20	16	8	5	11	6	1	12	874
1982	0.91	0.95	0.42	7	10	19	16	9	5	13	7	2	12	864
1983	0.90	1.39	0.45	6	9	18	18	9	6	13	9	2	11	870
1984	0.89	1.29	0.43	6	9	18	17	8	6	15	8	2	11	874
1985	0.90	1.56	0.40	5	9	17	16	8	6	18	8	2	11	891
1986	0.91	1.99	0.39	4	9	16	18	8	6	20	7	2	12	891
1987	0.91	2.34	0.39	4	10	15	17	8	6	19	8	2	11	893
1988	0.91	2.12	0.38	4	10	15	17	7	6	19	8	3	11	915
1989	0.92	2.51	0.37	4	10	15	17	8	6	18	7	3	11	915
1990	0.93	2.56	0.38	5	10	15	16	8	7	17	8	3	12	913
1991	0.93	2.93	0.39	5	9	15	16	8	8	17	8	3	12	904
1992	0.92	3.38	0.39	5	9	14	17	8	8	17	9	3	11	911
1993	0.90	3.78	0.36	5	9	13	17	7	7	18	9	3	11	906
1994	0.89	3.84	0.37	5	9	13	18	7	7	17	11	3	9	898
1995	0.89	4.59	0.38	5	9	12	16	7	8	17	13	3	9	891
1996	0.89	5.74	0.39	6	8	13	16	7	8	17	14	3	9	867
1997	0.90	7.30	0.39	6	8	12	16	7	9	17	15	2	8	865
1998	0.91	9.07	0.37	5	7	12	17	7	9	17	15	3	8	852
1999	0.94	11.29	0.38	5	6	12	17	6	9	16	19	3	7	856
2000	0.94	13.36	0.44	5	5	9	14	5	11	14	28	3	6	794
2001	0.93	11.53	0.41	5	5	9	16	6	12	17	21	2	6	846
2002	0.93	9.98	0.40	5	6	11	18	6	12	18	17	1	6	876
2003	0.93	9.82	0.40	6	5	10	19	6	11	17	17	1	6	914
2004	0.92	11.61	0.40	6	6	11	19	6	12	17	16	1	6	923
2005	0.92	12.46	0.41	7	6	12	18	6	11	17	16	1	6	902
2006	0.92	13.33	0.41	7	6	14	17	6	11	17	15	1	6	884
2007	0.93	14.91	0.40	8	7	14	17	6	11	15	15	2	6	903
2008	0.93	12.02	0.39	10	7	15	15	6	11	14	15	1	6	892
2009	0.94	9.58	0.41	8	7	15	16	6	12	13	15	2	7	906
2010	0.94	11.83	0.41	8	7	15	17	6	11	13	16	1	6	918
2011	0.94	13.19	0.42	8	7	15	17	6	10	13	17	1	6	928
2012	0.94	14.03	0.39	8	7	15	17	6	11	13	16	1	6	917
2013	0.94	16.67	0.38	8	7	16	18	6	10	14	15	1	6	930
2014	0.94	19.28	0.37	8	7	16	18	6	10	14	16	1	6	923
2015	0.94	20.37	0.35	7	6	16	15	6	12	15	15	1	6	917
Mean	0.92	7.90	0.40	6	8	14	17	7	9	16	13	2	8	892

Table 2.2: Summary statistics of the firm-specific network and implications in asset pricing. For each sample firm i, the number of interconnections of firm i in $\hat{\Omega}_t$ are added, representing the ith firm's degree of the firm-specific connections at year t, denoted as D_{it}. For each year, the total number of interconnections in $\hat{\Omega}_t$ compared to the total number of possible interconnections $p(p-1)$ is calculated, describing the aggregate degree of the firm-specific connections or the degree of the marketwide-specific connections, denoted as M_t. The aggregate degree of the firm-specific connections (M_t), the minimum, 5%, 25%, median, 75%, 95%, and maximum of the firm-level degree of the firm-specific connections D_{it} are reported in the table, respectively. The cross-sectional average pairwise-correlation of the residuals (partial correlation) and variance of the residuals from the regressions of stock excess returns on the Fama and French five factor model are presented in the last two columns of the table, respectively.

	Firm-specific connections								Partial	Residual
Year	M_t	min	5%	25%	median	75%	95%	max	correlation	variance
1980	0.0343	11	23	30	34	39	46	63	0.0582	3.950
1981	0.0345	15	23	29.75	34	39	47	62	0.0545	3.430
1982	0.0346	12	23	30	34	39	46	55	0.0556	3.894
1983	0.0338	10	22	29	34	39	46	59	0.0543	3.547
1984	0.0343	15	22	29	34	39	48	61	0.0536	3.225
1985	0.0339	16	23	29	34	39	47	57	0.0536	2.691
1986	0.0336	14	21	28	33	39	46	58	0.0543	3.363
1987	0.0334	5	21	29	33	38	45	69	0.0688	4.499
1988	0.0332	12	21	28	33	38	46	55	0.0547	2.752
1989	0.0335	14	22	28	33	39	46	54	0.0542	2.520
1990	0.0336	11	22	29	33	38	46	58	0.0565	3.912
1991	0.0342	10	21	29	34	39	47	60	0.0546	4.003
1992	0.0339	12	22	29	33	39	46	56	0.0551	3.693
1993	0.0326	13	22	28	32	37	43	56	0.0557	3.490
1994	0.0334	13	23	29	33	38	45	59	0.0549	3.190
1995	0.0322	8	22	28	32	36	43	56	0.0556	3.211
1996	0.0329	6	22	28	33	38	44.05	56	0.0562	4.042
1997	0.0322	8	20	27	32	37	45	57	0.0564	4.266
1998	0.0326	10	20	28	33	37.25	44	55	0.0626	6.354
1999	0.0313	9	21	27	31	35.25	42	60	0.0600	8.378
2000	0.0292	4	18	25	29	34	40	50	0.0650	16.168
2001	0.0277	10	17	23	28	32	39	54	0.0712	8.738
2002	0.0268	10	17	22	27	31	37	48	0.0706	6.750
2003	0.0288	4	17	24	29	34	40	51	0.0608	3.403
2004	0.0295	8	19	25	29	34	41	56	0.0608	2.583
2005	0.0290	7	18	24	29	33	41	53	0.0631	2.301
2006	0.0298	3	18	25	30	35	42	53	0.0647	2.482
2007	0.0298	4	18	25	30	34	41	59	0.0724	2.858
2008	0.0280	6	17	23	28	32	39	52	0.0826	8.909
2009	0.0274	5	16	23	28	32	38	52	0.0773	6.475
2010	0.0288	7	16	24	29	34	42	50	0.0620	2.428
2011	0.0281	5	16	24	28	33	40	54	0.0667	2.717
2012	0.0282	4	17	24	28	33	39	52	0.0629	2.586
2013	0.0279	8	16	23	28	32	40.05	50	0.0603	2.108
2014	0.0264	6	16	22	26	31	38	48	0.0692	2.759
2015	0.0255	8	14	20.75	25	30	37	50	0.0739	2.599
Mean	0.0311	8.97	19.61	26.35	30.92	35.74	42.84	55.50	0.0615	4.285

Table 2.3: Change points in the aggregate degree of the firm-specific network and the cross-sectional average partial correlation.
We apply the "Binary Segmentation" method proposed by [50] to detect the change points in the mean of the aggregate degree of the firm-specific connections (M_t) over the whole sample period from 1980 to 2015 and the change points in the cross-sectional average of the partial correlations among 1,000 sample firms (ϱ_t) given the Fama and French five factors. To implement the binary segmentation method, first, we multiply M_t by $1,000 \times 999$ to get the possible total number of conditional interconnections for the 1,000 sample firms each year from 1980 to 2015, denoted as $M_{T,t}$. We then apply the function "cpt.mean" in the R package "changepoint" ([51]) to detect the significant changes in the mean of $M_{T,t}$ over the sample period. Similarly, for the cross-sectional partial correlation of the 1,000 sample firms, we multiply $\bar{\varrho}_t$ by $1,000 \times 999$ to get the total value of the partial correlations among the 1,000 sample firms and then apply the binary segmentation method to detect the change years in the mean of the aggregate value of the partial correlations. The first five significant change years in the mean of the aggregate number of the firm-specific connections and the aggregate value of the partial correlations among the 1,000 sample firms are presented in the table by the order of significance, respectively.

Significance order	1st	2nd	3rd	4th	5th
M_t	1999	1992	2012	2002	2007
$\bar{\varrho}_t$	1999	2006	2009	2002	1997

Table 2.4: Effects of style, industry, and location on the firm-specific network. This table reports the common effect among firms (μ_t), the effects of style (β_t^S), industry (β_t^I), and headquarter location (β_t^L) on the firm-specific connections from the Bayesian nonparametic model (4a) to (4c) each year over the sample period from 1980 to 2015, respectively, with their corresponding 95% confidence intervals (CI).

		CI of μ_t			CI of β_t^S			CI of β_t^I			CI of β_t^L	
Year	μ_t	Lower	Upper	β_t^S	Lower	Upper	β_t^I	Lower	Upper	β_t^L	Lower	Upper
1980	-3.484	-3.501	-3.467	0.392	0.350	0.431	0.808	0.774	0.851	-0.009	-0.103	0.066
1981	-3.460	-3.475	-3.444	0.321	0.277	0.365	0.697	0.659	0.732	-0.031	-0.098	0.044
1982	-3.457	-3.473	-3.432	0.281	0.246	0.318	0.686	0.659	0.716	-0.068	-0.140	-0.017
1983	-3.488	-3.506	-3.471	0.289	0.259	0.319	0.678	0.651	0.706	-0.054	-0.112	0.017
1984	-3.468	-3.481	-3.453	0.309	0.270	0.341	0.612	0.580	0.646	-0.004	-0.061	0.051
1985	-3.491	-3.506	-3.479	0.338	0.306	0.375	0.632	0.605	0.663	0.032	-0.021	0.079
1986	-3.531	-3.549	-3.511	0.393	0.364	0.429	0.726	0.694	0.758	0.034	-0.021	0.079
1987	-3.544	-3.558	-3.529	0.428	0.399	0.461	0.707	0.680	0.736	0.010	-0.045	0.062
1988	-3.527	-3.542	-3.511	0.416	0.385	0.460	0.617	0.584	0.648	-0.001	-0.056	0.046
1989	-3.519	-3.536	-3.500	0.410	0.379	0.444	0.668	0.639	0.700	0.022	-0.035	0.079
1990	-3.543	-3.558	-3.527	0.395	0.369	0.422	0.782	0.752	0.809	0.044	-0.010	0.104
1991	-3.523	-3.538	-3.505	0.351	0.321	0.374	0.834	0.807	0.865	0.027	-0.020	0.077
1992	-3.547	-3.562	-3.528	0.332	0.303	0.367	0.916	0.889	0.946	-0.004	-0.049	0.043
1993	-3.614	-3.630	-3.596	0.339	0.298	0.377	1.035	1.010	1.068	-0.018	-0.067	0.029
1994	-3.604	-3.618	-3.588	0.328	0.298	0.361	1.083	1.057	1.114	-0.019	-0.086	0.029
1995	-3.651	-3.667	-3.634	0.347	0.317	0.380	1.136	1.112	1.165	-0.009	-0.067	0.045
1996	-3.664	-3.676	-3.642	0.398	0.369	0.427	1.201	1.174	1.232	0.011	-0.040	0.065
1997	-3.693	-3.711	-3.671	0.429	0.393	0.458	1.201	1.173	1.227	0.020	-0.046	0.085
1998	-3.680	-3.698	-3.663	0.444	0.412	0.473	1.178	1.150	1.203	0.011	-0.041	0.064
1999	-3.768	-3.784	-3.749	0.459	0.423	0.489	1.309	1.284	1.335	0.015	-0.033	0.068
2000	-3.990	-4.008	-3.967	0.455	0.426	0.485	1.584	1.550	1.612	0.019	-0.034	0.063
2001	-4.086	-4.105	-4.061	0.442	0.407	0.469	1.764	1.730	1.787	-0.010	-0.056	0.039
2002	-4.085	-4.105	-4.064	0.437	0.399	0.475	1.709	1.682	1.732	-0.037	-0.096	0.021
2003	-3.954	-3.967	-3.940	0.419	0.383	0.448	1.563	1.542	1.587	-0.010	-0.071	0.049
2004	-3.901	-3.918	-3.884	0.385	0.357	0.411	1.513	1.488	1.539	0.022	-0.034	0.072
2005	-3.884	-3.900	-3.864	0.350	0.317	0.385	1.472	1.447	1.497	-0.004	-0.052	0.049
2006	-3.821	-3.838	-3.804	0.346	0.312	0.383	1.389	1.362	1.420	-0.034	-0.087	0.025
2007	-3.843	-3.862	-3.824	0.387	0.350	0.426	1.444	1.415	1.472	-0.004	-0.055	0.057
2008	-3.975	-3.997	-3.959	0.413	0.377	0.446	1.645	1.619	1.669	0.038	-0.020	0.102
2009	-4.023	-4.039	-4.000	0.404	0.370	0.442	1.725	1.698	1.752	0.036	-0.014	0.089
2010	-3.967	-3.985	-3.949	0.409	0.369	0.444	1.688	1.664	1.719	0.025	-0.037	0.080
2011	-3.985	-4.002	-3.964	0.423	0.391	0.456	1.657	1.628	1.689	0.046	-0.015	0.095
2012	-3.944	-3.967	-3.925	0.405	0.376	0.436	1.563	1.534	1.590	0.054	-0.008	0.122
2013	-3.940	-3.957	-3.920	0.406	0.372	0.438	1.510	1.479	1.536	0.008	-0.055	0.067
2014	-4.066	-4.092	-4.049	0.451	0.411	0.504	1.667	1.639	1.699	-0.038	-0.113	0.025
2015	-4.029	-4.050	-4.008	0.458	0.415	0.510	1.707	1.669	1.739	-0.006	-0.084	0.088
Mean	-3.743	-3.760	-3.724	0.389	0.355	0.423	1.197	1.169	1.227	0.003	-0.055	0.060

Table 2.5: Firm clustering based on the firm-specific network.

Each year, from 1980 to 2015, we group the 1,000 sample firms into different communities based on $\hat{\Omega}_t$ by the fast greedy community detection algorithm ([43]). The number of communities, the modularity of the communities, and the number of firms within each community are reported in the table, respectively, for each year over the sample period.

Year	Community Number	Modularity	Number of firms within each community					
1980	5	0.218	381	364	137	114	4	
1981	4	0.208	372	344	261	23		
1982	4	0.193	464	408	112	16		
1983	5	0.196	397	342	232	18	11	
1984	4	0.171	410	400	179	11		
1985	5	0.167	363	361	231	37	8	
1986	5	0.192	430	421	145	2	2	
1987	4	0.262	498	409	91	2		
1988	5	0.221	518	374	95	7	6	
1989	6	0.199	460	359	158	18	3	2
1990	5	0.199	390	351	239	11	9	
1991	4	0.182	406	358	233	3		
1992	6	0.190	363	302	290	32	7	6
1993	5	0.219	398	298	222	65	17	
1994	6	0.203	382	310	234	61	8	5
1995	4	0.230	389	345	176	90		
1996	5	0.210	437	348	200	8	7	
1997	5	0.209	366	363	195	61	15	
1998	4	0.210	436	386	150	28		
1999	5	0.259	380	350	224	29	17	
2000	4	0.304	515	417	65	3		
2001	4	0.359	318	310	208	164		
2002	4	0.320	383	259	253	105		
2003	6	0.268	349	336	270	31	8	6
2004	4	0.248	392	332	269	7		
2005	5	0.246	409	334	219	31	7	
2006	4	0.234	432	384	143	41		
2007	4	0.262	361	334	168	137		
2008	5	0.307	381	302	297	11	9	
2009	4	0.299	400	339	149	112		
2010	4	0.265	352	318	252	78		
2011	4	0.290	366	308	307	19		
2012	5	0.248	431	350	109	104	6	
2013	5	0.253	348	284	248	108	12	
2014	4	0.286	362	301	278	59		
2015	5	0.306	356	316	164	113	51	

Table 2.6: Portfolio diversification.

Each year, from 1980 to 2015, we randomly select stocks (ranging from 1 to 20) per community clustered from the firm-specific network with replacement and form an equal-weighted portfolio. We then compare the Sharpe ratio (SHP) of the resulting equal-weighted portfolio with the SHP of the U.S. value-weighted market index. This process is repeated 5,000 times. This table reports the simulation average of the time-series mean (Mean) and standard deviations (Std) of the resulting equal-weighted portfolio returns (Pret) over all 5,000 replications, the simulation mean and standard deviations of the SHP of the resulting portfolios over all 5,000 replications, the proportion that the resulting portfolio yields a higher SHP than the market index among all 5,000 replications (PRO SHP), and the lower and upper bounds of the 95% confidence interval (CI) for the SHP for both in-sample and out-of-sample analysis over the whole sample period (1980–2015) and two equally divided sub-sample periods (1980–1997 and 1998–2015), respectively. The SHP of the U.S. market index is 0.136 over the period of 1980–2015, 0.130 over the period of 1981–2015, 0.181 over the period of 1980–1997, 0.172 over the period of 1981–1997, 0.096 over the period of 1998–2015, and 0.085 over the period of 1999–2015.

	In sample							**Out of sample**						
selected number	Mean Pret	Std Pret	Mean SHP	Std SHP	PRO SHP	Lower CI	Upper CI	Mean Pret	Std Pret	Mean SHP	Std SHP	PRO SHP	Lower CI	Upper CI
Panel A: whole sample period 1980–2015														
1	1.339	6.760	0.143	0.031	0.589	0.081	0.203	1.121	6.923	0.111	0.032	0.261	0.048	0.174
2	1.341	6.016	0.161	0.025	0.843	0.114	0.212	1.112	6.179	0.122	0.026	0.371	0.072	0.172
3	1.340	5.734	0.169	0.022	0.933	0.126	0.212	1.112	5.904	0.128	0.022	0.454	0.085	0.171
4	1.337	5.597	0.172	0.019	0.972	0.135	0.211	1.112	5.775	0.131	0.019	0.503	0.093	0.169
5	1.339	5.508	0.175	0.018	0.988	0.141	0.210	1.113	5.691	0.133	0.018	0.557	0.099	0.168
6	1.339	5.451	0.177	0.016	0.995	0.146	0.209	1.113	5.635	0.134	0.016	0.594	0.102	0.165
7	1.339	5.406	0.179	0.015	0.997	0.149	0.209	1.115	5.592	0.135	0.015	0.629	0.106	0.164
8	1.339	5.375	0.180	0.014	0.999	0.153	0.207	1.111	5.557	0.136	0.014	0.644	0.108	0.164
9	1.339	5.353	0.180	0.014	0.999	0.154	0.208	1.114	5.538	0.137	0.013	0.675	0.110	0.163
10	1.338	5.329	0.181	0.013	1.000	0.157	0.207	1.113	5.517	0.137	0.013	0.694	0.111	0.163
11	1.339	5.314	0.182	0.012	1.000	0.157	0.206	1.113	5.500	0.137	0.012	0.713	0.114	0.161
12	1.340	5.300	0.182	0.012	1.000	0.159	0.205	1.115	5.489	0.138	0.012	0.737	0.116	0.161
13	1.339	5.291	0.183	0.011	1.000	0.161	0.205	1.114	5.478	0.138	0.011	0.750	0.115	0.160
14	1.339	5.277	0.183	0.011	1.000	0.162	0.205	1.113	5.464	0.138	0.011	0.770	0.117	0.160
15	1.339	5.272	0.183	0.011	1.000	0.163	0.204	1.112	5.459	0.138	0.011	0.775	0.117	0.159
16	1.340	5.263	0.184	0.010	1.000	0.164	0.204	1.114	5.449	0.139	0.010	0.791	0.118	0.159
17	1.339	5.255	0.184	0.010	1.000	0.164	0.203	1.113	5.443	0.139	0.010	0.800	0.119	0.159
18	1.339	5.248	0.184	0.010	1.000	0.165	0.204	1.113	5.440	0.139	0.010	0.800	0.120	0.158
19	1.338	5.246	0.184	0.009	1.000	0.166	0.203	1.112	5.435	0.139	0.009	0.813	0.120	0.157
20	1.338	5.239	0.184	0.009	1.000	0.166	0.202	1.114	5.428	0.139	0.009	0.825	0.121	0.158

	In sample							Out of sample						
selected number	Mean Pret	Std Pret	Mean SHP	Std SHP	PRO SHP	Lower CI	Upper CI	Mean Pret	Std Pret	Mean SHP	Std SHP	PRO SHP	Lower CI	Upper CI
Panel B: first sub-sample period 1980–1997														
1	1.513	6.050	0.155	0.042	0.268	0.074	0.242	1.316	6.149	0.124	0.046	0.148	0.034	0.213
2	1.515	5.450	0.173	0.034	0.399	0.106	0.240	1.316	5.544	0.137	0.036	0.165	0.067	0.207
3	1.510	5.231	0.179	0.029	0.470	0.122	0.235	1.319	5.333	0.143	0.030	0.173	0.085	0.201
4	1.512	5.126	0.183	0.026	0.530	0.133	0.234	1.317	5.222	0.146	0.028	0.167	0.091	0.202
5	1.514	5.056	0.186	0.023	0.580	0.141	0.232	1.313	5.155	0.147	0.025	0.152	0.098	0.196
6	1.512	5.012	0.187	0.021	0.607	0.144	0.229	1.315	5.110	0.148	0.022	0.144	0.105	0.193
7	1.514	4.979	0.188	0.020	0.651	0.150	0.227	1.316	5.074	0.150	0.021	0.139	0.109	0.191
8	1.512	4.952	0.189	0.019	0.674	0.152	0.227	1.318	5.048	0.151	0.020	0.145	0.113	0.190
9	1.513	4.933	0.190	0.018	0.703	0.155	0.226	1.317	5.029	0.151	0.019	0.125	0.115	0.188
10	1.512	4.916	0.191	0.017	0.722	0.158	0.224	1.317	5.014	0.152	0.018	0.127	0.116	0.188
11	1.514	4.905	0.191	0.016	0.740	0.160	0.224	1.317	5.001	0.152	0.017	0.117	0.119	0.185
12	1.512	4.893	0.191	0.016	0.752	0.161	0.222	1.316	4.992	0.152	0.017	0.115	0.119	0.186
13	1.513	4.885	0.192	0.015	0.775	0.164	0.221	1.317	4.982	0.153	0.016	0.108	0.121	0.184
14	1.512	4.877	0.192	0.015	0.776	0.164	0.221	1.317	4.975	0.153	0.015	0.104	0.124	0.182
15	1.513	4.871	0.192	0.014	0.801	0.165	0.220	1.316	4.969	0.153	0.015	0.093	0.124	0.181
16	1.513	4.865	0.193	0.014	0.805	0.166	0.219	1.316	4.961	0.153	0.014	0.089	0.125	0.181
17	1.513	4.860	0.193	0.013	0.822	0.168	0.219	1.315	4.957	0.153	0.014	0.080	0.126	0.180
18	1.514	4.858	0.193	0.013	0.845	0.168	0.218	1.315	4.953	0.153	0.013	0.078	0.127	0.179
19	1.514	4.852	0.193	0.012	0.844	0.169	0.218	1.316	4.948	0.154	0.013	0.068	0.128	0.179
20	1.513	4.847	0.193	0.012	0.850	0.170	0.217	1.317	4.943	0.154	0.013	0.073	0.129	0.179
Panel C: second sub-sample period 1998–2015														
1	1.167	7.384	0.136	0.045	0.808	0.047	0.226	0.926	7.515	0.103	0.047	0.647	0.010	0.196
2	1.162	6.520	0.153	0.036	0.948	0.082	0.226	0.924	6.663	0.116	0.037	0.800	0.044	0.189
3	1.165	6.203	0.161	0.031	0.982	0.100	0.222	0.931	6.367	0.122	0.032	0.876	0.059	0.185
4	1.166	6.033	0.165	0.028	0.994	0.111	0.222	0.923	6.209	0.124	0.028	0.920	0.069	0.178
5	1.166	5.934	0.168	0.026	0.999	0.120	0.218	0.926	6.105	0.126	0.026	0.946	0.077	0.178
6	1.164	5.860	0.170	0.024	0.999	0.125	0.218	0.923	6.039	0.127	0.023	0.966	0.082	0.173
7	1.168	5.811	0.172	0.022	1.000	0.128	0.216	0.926	5.988	0.129	0.022	0.975	0.085	0.170
8	1.165	5.767	0.173	0.021	1.000	0.133	0.213	0.924	5.958	0.129	0.021	0.984	0.088	0.170
9	1.166	5.741	0.174	0.020	1.000	0.135	0.214	0.927	5.931	0.130	0.020	0.994	0.093	0.168
10	1.165	5.716	0.174	0.019	1.000	0.138	0.212	0.923	5.908	0.130	0.019	0.991	0.092	0.167
11	1.164	5.701	0.174	0.018	1.000	0.139	0.210	0.923	5.889	0.130	0.018	0.994	0.095	0.166
12	1.168	5.684	0.176	0.017	1.000	0.143	0.210	0.927	5.871	0.131	0.017	0.996	0.097	0.165
13	1.167	5.668	0.176	0.017	1.000	0.143	0.209	0.925	5.864	0.131	0.017	0.998	0.099	0.164
14	1.166	5.655	0.176	0.016	1.000	0.144	0.209	0.925	5.848	0.131	0.016	0.997	0.100	0.163
15	1.168	5.644	0.177	0.016	1.000	0.147	0.209	0.927	5.836	0.132	0.016	0.999	0.102	0.163
16	1.166	5.634	0.177	0.015	1.000	0.147	0.207	0.925	5.832	0.132	0.015	1.000	0.103	0.161
17	1.166	5.631	0.177	0.015	1.000	0.148	0.206	0.927	5.822	0.132	0.015	0.999	0.103	0.162
18	1.165	5.620	0.177	0.015	1.000	0.149	0.206	0.923	5.815	0.132	0.014	0.999	0.103	0.160
19	1.166	5.614	0.177	0.014	1.000	0.150	0.205	0.925	5.808	0.132	0.014	0.999	0.105	0.160
20	1.165	5.607	0.178	0.014	1.000	0.151	0.204	0.926	5.802	0.133	0.014	1.000	0.106	0.160

Table 2.7: Summary statistics of the firm-specific connections sorted portfolios.

Each year, from 1980 to 2015, we sort the 1,000 sample stocks by D_{it} and form quintile equal-weighted portfolios over that year. Portfolio 1 has the lowest degree of the firm-specific connections, while portfolio 5 has the largest degree of the firm-specific connections. The out-of-sample portfolios are constructed using the same stocks but monthly returns over the next year. For each industry sector, energy (ENG), materials (MTR), industrials (IDL), consumer discretionary (CDS), consumer staples (CST), health care (HCR), financials (FNL), information technology (IT), telecommunication service (TEL), and utility (UTL), the percentage of firms from 1,000 sample firms in each portfolio belonging to the industry sector is reported in Panel A. The time-series mean and standard deviation of the portfolio returns and the portfolio Sharpe ratios are reported in Panel B for both in-sample and out-of-sample analyses, respectively.

Panel A: Percentage of firms within each industry sector

	ENG	MTR	IDL	CDS	CST	HCR	FNL	IT	TEL	UTL
P1 Sparse	37.452	21.682	16.411	22.855	28.857	20.646	18.901	14.110	28.053	27.156
2	23.848	16.374	16.386	18.992	21.117	19.244	18.039	17.435	23.651	26.183
3	16.813	16.054	15.503	17.710	18.704	18.535	17.758	19.280	16.035	18.205
4	12.862	19.576	20.415	19.018	16.442	19.967	19.558	22.373	13.931	16.077
P5 Dense	9.025	26.314	31.285	21.425	14.880	21.607	25.745	26.802	18.330	12.378

Panel B: Portfolio performance

	P1	P2	P3	P4	P5
In sample					
Return Mean	0.012	0.014	0.014	0.016	0.017
Return Std	0.043	0.049	0.052	0.055	0.060
Portfolio SHP	0.183	0.202	0.203	0.220	0.215
Out of sample					
Return Mean	0.010	0.011	0.011	0.010	0.011
Return Std	0.048	0.049	0.053	0.054	0.058
Portfolio SHP	0.136	0.149	0.135	0.126	0.123

CHAPTER 3

HOLDING-LINKED NETWORK IN MUTUAL FUNDS AND THE PREDICTABILITY OF FUND PERFORMANCE

3.1 Introduction

Managers' portfolio decisions are vital to funds' performance and flows. A variety of explanations for the portfolio holdings are proposed, such as skill, private signals, agency problem, and irrationality. Regardless of numerous potential drivers, the resulting portfolio reflects all underlying drivers. Managers who make similar portfolio decisions are likely to have similar drivers, and then are likely to make similar decisions in the near future. Therefore, funds' past portfolio decisions might be able to predict their similar counterparts' subsequent portfolio choices, which could manifest in funds' performance and flows. In this paper, we quantify the fund-to-fund similarity in holdings, resulting in a holding-linked network of funds. We then investigate the cross-fund predictability of fund performance and flows through the network, which captures all underlying drivers of funds' portfolios and demonstrates significant and persistent predictability.

To measure the fund-to-fund similarity in holdings, we use vector representations of funds' holdings to define locations for funds in stock space. The similarity between every two funds' portfolios is measured by the distance between their locations in their stock-space. This measurement of similarity includes not only the assets in holdings, but also the corresponding holding levels. Changes in holding assets or levels vary similarities. Therefore, the measurement provides a continuous and dynamic representation of the pairwise similarity of any two funds.

Each month over the whole sample period, from December 1992 to December 2015, we compute the fund-to-fund similarity in holdings for 3,569 unique funds in our sample. There is a considerable cross-sectional heterogeneity in the similarities per month, and the difference persists over time. For example, the maximum similarity is 1, indicating identical holdings, while the minimum is 0, implying orthogonal holdings. We then measure a fund's similarity in holdings with others by the average fund-to-fund similarity between the fund and the others. We find that funds with high similarity in holdings with others tend to have low return relative to their contrarian counterparts, which is consistent with previous literature (e.g., [58], [59], and [60]). In addition, some characteristics of funds are positively related with their similarity in holdings with other funds, such as the number of held stocks, total net asset under management (TNA), and fund age, while some other characteristics show negative relations, such as expense ratio, turnover, and manager tenure.

To examine the predictability of the holding similarity for fund performance, we sort all funds monthly into quintiles by their similarities and $alpha$s and form equally weighted portfolios the following month. The $alpha$s are estimated by regressing monthly excess returns on monthly Fama-French and Carhart factors over three-year rolling windows. We find that a long-short equity strategy, that is the portfolio constructed from the highest similarity and the lowest alpha portfolio minus the lowest similarity and the lowest alpha portfolio, yields significant annual alpha of 4.458% with t-statistics 4.926, indicating the significant predictability of funds' holding similarity for fund performance.

It is worth noting that the fund-to-fund similarity quantifies a holding link between funds, describing a holding-linked network. To explore the effectiveness of the holding-linked network in the cross-fund predictability of fund performance and flows, we propose a network based regression with the network structure taken into consideration. We find that funds experience greater improvement in performance the following month, measured by characteristic selectivity (CS) proposed by [61], if their holding-linked (non-linked) neighbors in the network experienced higher (lower) performance previous month. In addition, the larger the similarity between the fund and its linked neighbors, the greater the improvement is. Similar findings are obtained for the cross-fund prediction of fund flows through the holding-linked network.

To corroborate the significant cross-fund predictability of fund performance and flows through the holding-linked network, we add two other well-known predictors in our regression, i.e., return gap ([59]) and R^2 ([62]). The significance of predictability remains. In addition, we find that R^2 becomes positive and significant in predicting characteristic timing (CT, [61]) in the presence of the holding-linked network, which is insignificant in [62]. Lower R^2 indicates greater selectivity. That is, on average, funds with low selectivity tend to have high market timing ability.

We next examine the persistence of the cross-fund predictability of fund performance and flows though the holding-linked network. Specifically, we investigate the impact through funds' holding-linked network at previous 1, 3, 6, 9, and 12 months. We find that the previous one-month and three-month holding-linked network demonstrates significant effect in the cross-sectional predictability of fund performance measured by CS, and the previous three-month and six-month network shows significant effect in the prediction of CT. For CS, the impact through the previous one-month network is significantly positive, which is consistent with our previous results, while the impact through the previous three-month network is significant negative, which indicating the persistence of fund performance CS, less than three months. In addition, the impact through the network in predicting fund flows is persistent at least for one year.

Some studies use fund holding data to measure fund activity, then to investigate fund performance or flows. For example, [63] find that funds whose stock holdings are related to company-specific information that differs from analysts' expectations exhibit better performance; [64] measures manager skill using the stock holdings, which can predict fund performance. However, the cross-fund predictability through network, has, to our knowledge, not been documented in prior literature. The proposed approach takes all information from funds jointly and simultaneously into consideration through network in examining the cross-fund predictability and does not utilize any factor model.

Overall, our analysis shows significant and persistent cross-fund predictability of fund performance and flows through the holding-linked network, which can be easily constructed from the fund-to-fund similarity in holdings.

The remainder of the paper is organized as follows. Section 3.2 describes the data and sample selection procedure, and presents the method for the measurement of fund's similarity and holding links. In Section 3.3 and Section 3.4, we empirically examine whether the fund-to-fund holding links predicts fund's performance and flows, using various approaches and measurements of fund performance, followed by the comparison with other existing predictors. In Section 3.5, we investigate the persistence of the fund holding links in the prediction of fund's performance and flows. Section 3.6 examines the determinants of funds' holding similarities. Section 3.7 concludes the paper.

3.2 Data and Measurement

We start by describing our data and sample selection procedure. We present the method for the measurements of fund-to-fund's similarity and use the data to estimate them, resulting in the holding-linked network of funds.

3.2.1 Data and Sample Selection

Our data sample spans the period December 1992 – December 2015.[1] We use two major datasets for most of our main analysis. We begin with the Center for Research on Security Prices (CRSP) Survivorship Bias Free Mutual Fund Database. The CRSP database provides us with mutual fund returns as well as a host of fund characteristics such as fund size, fund fee, fund age, turnover, etc. at fund share class level. The CRSP data starts in 1962, however, about 15% of the funds on CRSP report only annual returns before 1980 and there is selection bias between funds that report annually and funds that report monthly ([65]). After 1980, almost all funds report monthly returns. Following [64], we only keep domestic open-end diversified U.S. equity funds since data availability and holding data for those funds are most reliable.[2] We further exclude index funds and sector funds from our sample.[3] Since the reported objectives do not always indicate whether a fund portfolio is balanced, we also exclude fund-month observations in which the fund holds less than 80% of its total net asset under management (TNA) in stocks. For mutual funds that have multiple share classes, we aggregate different share classes of the same fund into one single observation since

[1]The manager tenure data computed from CRSP is available from December 1992.

[2]We identify fund styles based on the objective codes and on the disclosed asset composition. We exclude funds with Thompson Reuters Database Objective Codes: International, Municipal Bonds, Bond and Preferred, and Balanced. We include funds with the following Strategic Insight objectives: AGG, GMC, GRI, GRO, ING, or SCG. If a fund does not have any of the above Strategic Insight objectives, we select funds with the following Weisenberger objectives: GCI, IEQ, IFL, LTG, MCG, SCG, G, G-I, G-I-S, G-S, G-S-I, GS, I, I-G, I-G-S, I-S, I-S-G, S, S-G-I, S-I, or S-I-G. If a fund has neither a Weisenberger nor Strategic Insight objective, then we go to the Lipper objective codes and pick funds with the following objectives: CA, EI, EIEI, ELCC, G, GI, LCCE, LCGE, LCVE, LSE, MC, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, MR, S, SCCE, SCGE, SCVE, SESE, and SG. If none of these objectives are available and the funds have policy codes CS, Flex or I-S, then the fund will be included.

[3][66] states that CRSP does not provide any way to discriminate between active and passively managed funds'. Those two authors use identifier provided by Morningstar and further augment it by manually checking to separate passively managed fund from actively managed ones. Limited by not having access to Morningstar database. We identify index funds by searching the funds' names with the keywords "index", "inde", "indxv", "inx", "idx", "dow", "jones", "ishare", "s&p", "S&P", "500", "WILSHIRE", "RUSSELL", "RUSS", and "MSCI".

they have the same stock holdings.[4]

To address the possibility of incubation bias commented by [67][5], we apply the following filters. We exclude observations for which the year of the observation is prior to the reported fund starting year as well as observations for which the names of the funds are missing in the CRSP database. Incubated funds also tend to be smaller, which motivates us to exclude funds that in the previous month had less than $5 million on assets under management or fewer than 10 stocks.

Next, we merge the CRSP mutual fund data with the Thompson and Reuters S12 stock holdings database using MFLINK. Thompson and Reuters S12 collect fund stock holdings data both from reports filed by mutual funds with the SEC and from voluntary reports generated by the funds.[6] We then link reported stock holdings to the CRSP stock database in order to calculate fund returns. Finally, we map funds to the names of their managers using information from CRSP to compute manger tenure.[7]

We end up with a sample of 3,596 unique funds over our whole sample period from December 1992 to December 2015. In our study, we scale fund dollar flows by beginning-of-month assets under management to capture the change in size due to net money-flows (e.g., [68] and [69]). That is, flow of fund i at month t is defined by $\text{Flow}_t^i = (\text{TNA}_t^i - \text{TNA}_{t-1}^i \times \text{ret}_{t-1}^i)/\text{TNA}_{t-1}^i$, where ret_{t-1}^i is fund i's net return at month $t-1$. In addition, all the variables in our study are winsorized by the top and bottom 0.5% tails.

[Table 3.1 inserted here]

Table 3.1 shows summary statistics for our sample funds. There are 1,804 unique funds per month on average over the whole sample period and each fund holds stocks ranging from 11 to 1,940 with an average, 143. The average fund has $904.159 millions in assets, expenses of 1.200%,

[4] We sum the TNA of share classes. For the qualitative attributes of funds (e.g., names, objectives, and year of origination), we retain the observation of the oldest fund. For the other attributes of funds (e.g., returns, expenses, loads), we take the weighted average, where the weights are the lagged TNAs of each share class.

[5] Bias can arise when fund families incubate several private funds and then only make public the track record of the surviving incubated funds, not the terminated funds.

[6] The Thomson and Reuters S12 does not require that funds survive. During our sample period before May 2004, funds are required by law to disclose their holdings semiannually, although about 49% disclose quarterly. After May 2004, funds are required by law to disclose their holdings every quarter.

[7] Kacperzyk, Nieuwerburgh and Veldkamp (2014) also use Morningstar, Nelson's Directory of Investment Managers, Zoominfo, and Zabasearch. But since we do not have access to those databases we based our analysis solely on CRSP.

and turnover of 80.843%. In addition, the average fund in the sample is 11.151 years old, and the average manager has 6.058 years of experience. The average fund received an inflow of 0.461%, while the median fund received an outflow of 0.119%.

3.2.2 Measurement of Similarity

Each fund's portfolio can be thought of as having a location in stock-space that is determined by its weights on stocks. As the weights change, so does the location. In this paper, we measure the similarity between every two funds by the "distance" between the funds' locations in their stock-space.

In particular, let $n_t^{i,j}$ represent the number of stocks held by either fund i or fund j at month t, then the stock-space of funds i and j at month t is $n_t^{i,j}$-dimension. $H_{i,t}^{i,j} = (\omega_{1,t}^i, \omega_{2,t}^i, \cdots, \omega_{n_t^{i,j},t}^i)' \in \mathbb{R}^{n_t^{i,j}}$ describes the location of fund i in the stock-space spanned by funds i and j, where $\omega_{k,t}^i$ is the fraction of fund i's total assets held in stock k at month t.[8] If stock k is not in fund i's portfolio at month t, $\omega_{k,t}^i = 0$. Thus $\sum_{k=1}^{n_t^{i,j}} \omega_{k,t}^i = 1$ for any fund i.

The *similarity* between funds i and j at month t is measured by the cosine of the angle between their locations in the stock-space, that is,

$$Similarity_t^{i,j} = \cos(H_{i,t}^{i,j}, H_{j,t}^{i,j}) = \frac{H_{i,t}^{i,j} \cdot H_{j,t}^{i,j}}{\|H_{i,t}^{i,j}\|_2 \|H_{j,t}^{i,j}\|}, \tag{3.1}$$

where $\|\cdot\|_2$ is the Euclidean norm and $Similarity_t^{i,j} \in [0, 1]$.[9] If two funds have identical holdings of stocks, i.e., the same stocks and the corresponding weights, the angle between the locations of these two funds in their stock-space would equal zero, thus the similarity between these two funds is 1. If two funds have mutually exclusive holdings of stocks, i.e., orthogonal holdings, the angle would equal 90-degree and the similarity is 0. The larger the similarity is, the closer the holdings of funds are.

This approach to measure the similarities is widely used in a variety of areas, such as financial economics, text analysis, biostatistics, and physics (e.g., [70], [71], [72], [73], and [74]). As

[8] $\omega_{k,t}^i = (\text{shares}_{i,k,t} \times p_{k,t})/(\sum_k \text{shares}_{i,k,t} \times p_{k,t})$,where $\text{shares}_{i,k,t}$ is the number of shares of stock k held by fund i at the month t, and $p_{k,t}$ is the price of stock k at the month t.

[9] Since cosine has the nice property that it is 1.0 for identical vectors and 0.0 for orthogonal vectors.

explained in [74], the cosine of the angle between funds' vectors ($H_{i,t}^{i,j}$ and $H_{j,t}^{i,j}$) is more appropriate than the correlation coefficient between the two vectors because stocks' weights and changes in weights are constrained. Changes in weights sum to 0 and weights sum to 1. Therefore, any change, or any weight, is a linear combination of the others. Including all weights or changes in weights as independent observations overstate significance of correlation coefficients between the vectors.

Each month over the whole sample period, we calculate the fund-to-fund similarity in holdings for all sample funds through the measurement (3.1). Panel A of Table 3.2 summarizes the statistics of these similarities.

[Figures 3.1, 3.2, and Table 3.2 inserted here]

The similarity between funds is 0.096 on average with standard deviation 0.147 over the whole sample period. While the average similarity is close to zero, there is a large cross-sectional variation in the similarities between funds, ranging from 0 to 1. A more complete picture of the distribution of similarities between funds is presented in Figure 3.1, where the distribution appears to be strongly positive skewed. That is, a large proportion of funds hold very different portfolios. The differential persists over the whole sample period as presented in Figure 3.2. There is no obvious increasing trend in the cross-sectional mean and median similarity between funds over the whole sample period. The mean and median similarity is 0.099 and 0.029 over the earlier sub-sample period (December 1992 – December 2003) and is 0.095 and 0.031 over the recent sub-sample period (January 2004 – December 2015). In addition, the skewness of the distribution increases from 2.301 over the earlier sub-sample period to 2.312 over the recent sub-sample period. These results indicate the persistent heterogeneity in similarities between funds' portfolios, although funds become slightly more similar in their holdings over recent sub-sample period.

3.2.3 Holding Links of Mutual Funds

The fund-to-fund similarity measured in (3.1) quantifies a link between funds through their holdings — not only the holding assets, but also the holding levels of assets.

Let N_t represents the number of unique funds at month t. We define a "holding-linked network" among these N_t funds at month t by a matrix, $Similarity_t = (s_t^{i,j}) \in \mathbb{R}^{N_t \times N_t}$, where $s_t^{i,j} = s_t^{j,i} =$

$Similarity_t^{i,j}$ for $i \neq j$ and $s_t^{i,i} = 0$. Note that $Similarity_t$ is the adjacency matrix of the network and $s_t^{i,i}$ allows no self-linked. Then,

$$Similarity_t^i = \frac{1}{N_t - 1} \sum_{j=1}^{N_t} s_t^{i,j} \tag{3.2}$$

measures the similarity of fund i and all other funds in the market at month t.

To briefly examine the fund and manager's characteristics that relate to the fund's similarity measured in (3.2), we sort funds into quintiles based on $Similarity_t^i$ every month. The average net return, gross return, number of held stocks, total net assets (TNA), age, expenses, turnover, manager tenure, and flow of funds in each quintile over the whole sample period are presented in Panel B of Table 3.2.

As shown in the table, the number of held stocks, TNA, and fund age are positively related with funds' similarities, while fund net return, gross return, expense ratio, turnover, manager tenure, and flow are negatively related with funds' similarities in general. For example, funds in the top quintile have smaller average gross returns (1.045%) than the returns (1.114%) in the bottom quintile. The pattern persists after adjusting for the fees of funds, with 0.968% and 1.001% in the top and bottom quintiles, respectively. These results are consistent with previous literature that funds with similar holdings tend to perform poorly relative to their contrarian counterparts (e.g., [58], [59], and [60]). Another example, fund flow monotonically decreases from top quintile (0.474%) to bottom quintle (1.108%). More discussion is provided in Section 3.6. All together, the results here indicate the possible predictability of the holding similarity for fund performance and flows.

3.3 Fund Portfolio Performance Based on Sorting by Lagged Similarity and Alpha

In this section, we examine a strategy that predicts fund performance, measured by $alpha$, based on the fund's lagged holding *similarity* and *alpha*.

Each month t over the whole sample period, we sort funds which exist at both months t and $t-1$ into quintiles by their $Similarity_{t-1}^i$ defined in equation (3.2), and within each quintile we sort funds into quintiles by their $alpha_{t-1}^i$. For each fund, $alpha_{t-1}^i$ is estimated by regressing monthly excess fund returns (over the T-bill rate) on the monthly Fama-French and Carhart (FFC,

[8] and [7]) six factor returns —market factor, size factor, value factor, profit factor, investment factor, and momentum factor—over thirty-six months preceding month t.[10] We sort by $alpha^i_{t-1}$ because the persistence in fund's performance (e.g. [75], [76], and [62]). This procedure produces twenty-five (5×5) portfolios with an equal number of funds in each.[11]

Then for each month t over the period from 1993 to 2015, we calculate the average monthly excess returns (over the T-bill rate) of the funds that that are included in each cell of the 5×5 cells, which generate the corresponding portfolio returns. These portfolio returns are regressed on the FFC six factors over the twenty-three years (1993–2015). We present for each portfolio the regression $alpha$ (the regression intercept) and its t-statistic in Table 3.3, using robust standard errors following [77].

[Table 3.3 inserted here]

The $alpha$, shown in each $alpha$-quintile row in the Panel A of Table 3.3, exhibits an upward trend from left (low $Similarity^i_{t-1}$) to right (high $Similarity^i_{t-1}$), indicating high holding similarity with other funds are followed by high risk-adjusted abnormal return. In particular, in the row "All", besides the general increasing trend, all $alpha$s are significant and negative at least at the 5% level from left to right.

We test whether funds with high similarity in holdings with other funds significantly outperform funds with low similarity in risk-adjusted return, by estimating the $alpha$ of a hypothetical portfolio of a long position in the highest similarity quintile funds and a short position in the lowest similarity quintile funds over each *alpha* quintile. The results are presented in the rightmost column of Table 3.3 under "High-Low". The return from this strategy yields an significant annual $alpha_t$ of 4.458% (t=4.926) and 2.568% (t=2.742) for the smallest and second-smallest $alpha^i_{t-1}$ quintiles at the 1% level, respectively; and 0.943% (t=2.209) and 1.815% (t=2.472) for the highest and second-highest $alpha^i_{t-1}$ quintlies at the 5% level, respectively.

We replicate this analysis using fund gross return before accounting for expenses (see Panel B of Table 3.3). The results, presented in Panel B, show again that high-minus-low *similarity*

[10] Similar results are obtained when using FFC four factors, i.e., market factor, size factor, value factor, and momentum factor.

[11] Each month, the number of funds in each cell of the matrix may be slightly different because the number of funds in each month does not always divide exactly by twenty-five.

portfolio has significant annual $alpha_t$ of 4.196% (t=4.685) and 2.419% (t=2.593) for the smallest and second-smallest $alpha^i_{t-1}$ quintiles at 1% level, respectively.

To summarize, the results in Table 3.3 demonstrate significant predictability in fund performance through funds' holding similarity. Funds' risk-adjusted excess return is higher for funds with high cross-sectional similarity in holdings.

3.4 Cross-fund Prediction of Fund Performance and Flows through Holding-Linked Network

$Similarity_t$ measures a holding-linked network of funds. Each link in the network quantifies the similarity between the two funds through their holding assets and the corresponding holding levels. In this section, we empirically examine the cross-fund prediction of fund performance and flows through the holding-linked network.

3.4.1 Prediction of Fund Performance Measured by Characteristics-Based Excess Return

Following [62] in the study of predicting fund performance, we measure the fund performance by the two measures proposed by [61], i.e., (1) "Characteristic Selectivity" (CS) and (2) "Characteristic Timing" (CT). CT measures how portfolio managers time their portfolio weightings on the characteristics of stocks in their portfolios, and CS measures how managers can select stocks that outperform the average stock having the same characteristics. In our study, the monthly returns of CS and CT are expressed in percent points and are multiplied by 12 to annualize them.

The similarity in holdings between funds reflects similar underlying drivers of managers' portfolios. Managers who make similar portfolio decisions are likely to make similar decisions in the near future, then holding linked funds might be able to cross-sectionally predict each other's subsequent performance. Motivated from these, we estimate a model with CS^i_t as the dependent variable, predicted by its lag CS^i_{t-1}, the effects of fund's linked neighbors' lagged performance CS^j_{t-1}, and fund's lagged characteristics that may affect the performance:

$$\mathrm{CS}^i_t = \alpha_t + \beta_{0,t} CS^i_{t-1} + \beta_{1,t} \frac{1}{\sum_{j=1}^{N_{t-1}} I(s^{i,j}_{t-1} > 0)} \sum_{j=1}^{N_{t-1}} s^{i,j}_{t-1} CS^j_{t-1} + X^{'}_{i,t-1}\gamma_t + \epsilon_{i,t}, \tag{3.3}$$

where $X_{i,t-1}$ is a vector of lagged controls, including fund flow, log of fund's total net asset, square of the log of fund's total net asset, expense ratio, turnover, log of fund age, log of manager tenure, number of fund's holding stocks, and six different fund styles.[12] The performance CS^i_{t-1} is the standard autoregressive impact, and the coefficient $\beta_{0,t}$ describes the momentum effect. The quantity $(\sum_{j=1}^{N_{t-1}} I(s^{i,j}_{t-1} > 0))^{-1} \sum_{j=1}^{N_{t-1}} s^{i,j}_{t-1} CS^j_{t-1}$ is the weighted average impact from fund i's holding-linked neighbors. Its associated parameter $\beta_{1,t}$ describes the similarity network (or holding-linked network) effect. The weights are measured by the holding similarity between funds, i.e., $s^{i,j}_{t-1}$. Note that we multiply $(\sum_{j=1}^{N_{t-1}} I(s^{i,j}_{t-1} > 0))^{-1}$ in front of the network effect term, since the number of funds' holding-linked neighbors could be different across funds and over time.

A similar model is estimated with CT^i_t as the dependent variable. Model (3.3) is estimated by the Fama-MaccBeth procedure ([78]), following [8] and [79], with Newey-West standard errors using the optimal lag.[13] For the CS^i_t and CT^i_t regressions, we estimate all the coefficients in model (3.3) over 276 monthly test periods, 1993–2015. The hypothesis is that $\beta_1 > 0$ in the model (3.3). That is, funds having similar holdings indicate similar underlying drivers of portfolios, which enhance the possibility of making similar subsequent portfolio decisions, thus cross-sectionally predict subsequent performance.

[Table 3.4 inserted here]

Table 3.4, which presents the estimation results in columns (2) and (6), shows that the holding-linked funds at previous month have significant cross-fund predictability in CS with positive sign, but it does not predict CT. For CS, the mean network effect $\beta_{1,t}$ is 4.672% with t-statistics $t = 5.001$. That is, if the performance CS of a fund's linked neighbors increases 1% a month, the fund's CS rises 4.672% multiply the average similarity of its linked neighbors in the following month. The higher the similarity with its linked neighbors is, the more improvement of the performance is. For CT, the mean network effect $\beta_{1,t}$ is positive, 1.000%, but insignificantly different from zero ($t = 1.238$). However, results from Section 3.5 shows that the holding-linked network has significant lagged predictability for CT. More details are discussed in Section 3.5. Besides the network effect

[12] The fund styles are obtained from CRSP.

[13] The holding data might be available semiannually before 2004, thus there might exist autocorrelation in the estimated coefficient in model (3.4). We adjust the standard errors in computing the t-statistics. The optimal lag is selected by the automatic bandwidth selection method proposed in [22].

$\beta_{1,t}$, we find that the momentum effect $\beta_{0,t}$ is positive and significant, 0.678% with $t = 62.654$ for CS and 0.028% with $t = 1.695$ for CT, indicating the persistence of fund performance. Moreover, the smaller value of $\beta_{0,t}$ compared with $\beta_{1,t}$ might indicate the "size" effect from other similar funds.

Note that in model (3.3), the completely dissimilar funds are not taken into consideration, i.e., $s_{t-1}^{i,j} = 0$. However, those funds might also have impacts on the other funds' performance, for example, indirectly through the stocks in their portfolios. In order to take all funds and information in the holding-linked network into consideration, we consider the model:

$$\begin{aligned} \text{CS}_t^i &= \alpha_t + \beta_{0,t} CS_{t-1}^i \\ &+ \beta_{1,t} \frac{1}{\sum_{j=1}^{N_{t-1}} I(s_{t-1}^{i,j} > 0)} \sum_{j=1}^{N_{t-1}} s_{t-1}^{i,j} CS_{t-1}^j + \beta_{2,t} \frac{1}{\sum_{j=1}^{N_{t-1}} I(s_{t-1}^{i,j} = 0)} \sum_{j=1}^{N_{t-1}} I(s_{t-1}^{i,j} = 0) CS_{t-1}^j \\ &+ X_{i,t-1}^{'} \gamma_t + \epsilon_{i,t}, \end{aligned} \tag{3.4}$$

where $\beta_{0,t}$ and $\beta_{1,t}$ represent the momentum and holding-linked funds' effects, respectively, same as in model (3.3). The quantity $(\sum_{j=1}^{N_{t-1}} I(s_{t-1}^{i,j} = 0))^{-1} \sum_{j=1}^{N_{t-1}} I(s_{t-1}^{i,j} = 0) CS_{t-1}^j$ is the average impact from fund i's non-linked funds. The associated parameter $\beta_{2,t}$ in model (3.4) describes the complete-dissimilarity-network effect. If a fund's non-linked neighbors' performance is not related to its performance, we expect $\beta_{2,t} = 0$. However, if these non-linked funds have cross-sectional predictability in fund's performance, we expect $\beta_{2,t} < 0$. That is, lower performance in a fund's complete dissimilar neighbors indicates higher performance of the fund. A similar model is estimated with CT_t^i as the dependent variable.

The estimation results, reported in the columns (3) and (7) in Table 3.4, show that non-holding linked funds have significantly cross-fund predictability in CS with negative sign, in the presence of holding linked funds' impact. The mean of complete dissimilarity effect $\beta_{2,t}$ is -0.156% with $t = -3.481$. That is, the performance CS of a fund's no-linked funds decreases 1% a month on average, the fund's performance CS rises 0.156% the following month. In addition, the mean coefficient $\beta_{1,t}$ is positive and significant for CS. For CT, there is no evidence of cross-sectional predictability of CT through linked or non-linked funds, which is consistent the previous results from model (3.3).

To further investigate the predictability of holding-linked network, we consider both the linked and non-linked funds' impact together in the following model:

$$\mathrm{CS}_t^i = \alpha_t + \beta_{0,t} CS_{t-1}^i + \beta_{3,t} \frac{1}{\sum_{j=1}^{N_{t-1}} I(s_{t-1}^{i,j} \neq 1)} \sum_{j=1}^{N_{t-1}} (1 - s_{t-1}^{i,j}) CS_{t-1}^j + X_{i,t-1}^{'} \gamma_t + \epsilon_{i,t}, \quad (3.5)$$

where $1 - s^{i,j}$ describes the dissimilarity between funds i and j. If two funds are completely dissimilar, then $1 - s_t^{i,j} = 1$. The quantity $(\sum_{j=1}^{N_{t-1}} I(s_{t-1}^{i,j} \neq 1))^{-1} \sum_{j=1}^{N_{t-1}} (1 - s_{t-1}^{i,j}) CS_{t-1}^j$ is the weighted average impact from all funds in the market except fund i's identical holding linked funds, where the weights are measured by the dissimilarity between funds. The associated parameter $\beta_{3,t}$ in model (3.5) describes the dissimilarity-network effect. A similar model is estimated with CT_j^i as the dependent variable. We hypothesis that $\beta_{3,t} < 0$. That is, funds with dissimilar holdings indicates different underlying drivers of portfolios, which enhance the possibility of making different subsequent portfolio decisions, thus different performance. The estimation results, reported in columns (4) and (8) of Table 3.4, support our hypothesis.

As shown in the table, the cross-fund performance prediction is significant for CS through dissimilarity-network, but is insignificant for CT, consistent with similarity-network result. When CS_t^i is the dependent variable, the mean coefficient $\beta_{3,t}$ is -5.316% with $t = 4.796$. That is, if the performance CS of all funds, excluding fund i's identical holding funds, decreases 1% a month on average, the fund's performance CS rises 5.316% multiplied by the average dissimilarity of all funds, except its identical funds, the following month. The higher the dissimilarity with other funds is, the more the performance increases. And the higher the similarity with other funds is, the smaller the impact of the dissimilarity-network. When CT_t^i is the dependent variable, the mean coefficient $\beta_{3,t}$ is negative (-1.326%) as expected, but is insignificant.

3.4.2 Prediction of Fund Flow

Having documented the significant cross-sectional predictability of the holding-linked network in fund performance, this subsection examines the cross-fund predictability through the network for fund flows.

Similar models as (3.3), (3.4), and (3.5) are estimated with fund flow as the dependent variable, and the vector $X_{i,t}$ of lagged controls in each model includes the log of total net asset, square of the log of total net asset, expense ratio, turnover, log of fund age, log of manager tenure, number of fund's holding stocks, fund net return, and six different fund styles. Without confusion, our discussion of predicting fund flow refers to models (3.3), (3.4), and (3.5). The estimation method used in the previous subsection is applied, i.e., Fama-MaccBeth procedure with Newey-West standard errors using optimal lag over 276 monthly test periods, 1993–2015.

[Table 3.5 inserted here]

The estimated results, presented in Table 3.5, show that the holding-linked funds have significant cross-fund predictability in fund flows and the non-linked funds do not contribute to the prediction. Specifically, the effect of the holding-linked funds $\beta_{1,t}$ in model (3.3) is 8.252% on average with $t = 2.939$. That is, if the flows of a fund's linked neighbors increase 1% a month, the fund's flows rise 8.252% multiplied by the average similarity of its linked neighbors in the following month. The higher the similarity between the fund and its linked neighbors is, the more the inflow the fund experiences. In addition, the holding non-linked funds does not cross-sectionally predict fund flows. Although the mean of complete-dissimilarity network effect $\beta_{2,t}$ in model (3.4) is negative, -0.018%, it is insignificant. Furthermore, when the dissimilarity network effect is examined, the mean coefficient $\beta_{3,t}$ in model (3.5) is negative and significant, -13.481% with $t = -3.841$, which is consistent with the holding-linked funds' effect.

3.4.3 Comparison with Other Predictors

Recent studies propose a few measures of fund activity or skill, which are shown as predictors of fund performance. For example, [59] propose a predictor of fund performance which measures managerial skill. The skill is quantified by the fund manager's unobserved actions as measured by the return gap that is the difference between the reported fund return and the return on a portfolio that invests in the previously disclosed fund holdings. [62] propose a predictor of fund performance which measures managerial selectivity. The managerial selectivity is quantified by the fund's R^2, which is obtained from a regression of the fund's returns on a multifactor benchmark model and measures selectivity.

We reestimate models (3.3), (3.4), and (3.5), adding lagged return gap and logistic transformation of R^2 ([62], TR^2) as the explanatory variables in the models. Table 3.6, presented the estimation results, shows that cross-sectional predictability of the holding-linked network for fund performance and flow remains statistically significant in the presence of return gap and TR^2.

[Table 3.6 inserted here]

Specifically, for CS, the mean coefficient $\beta_{1,t}$ in model (3.3) is 4.743% with $t = 5.015$, which is even larger than the value without return gap and TR^2 as explanatory variables. However, return gap and TR^2 are insignificant. For CT, $\beta_{1,t}$ is insignificant in model (3.3) and (3.4), consistent with our previous results. But $\beta_{3,t}$ becomes negative and significant in model (3.5) for CT, -1.907% with $t = -1.774$.

It is interesting to note that when CT is the dependent variable, the effect of TR^2 is positive and significant in the presence of holding-linked network effect, 0.188% with $t = 2.340$, 0.185% with $t = 2.464$, and 0.192% with $t = 2.585$ in models (3.3), (3.4), and (3.5), respectively. TR^2 is insignificant in explaining CT in [62]. As explained in [62], lower TR^2 indicates greater selectivity. Thus, the positive and significant effect of TR^2 for CT implies that funds with great selectively have weak timing ability on average.

When fund flow is the dependent variable, the mean coefficient $\beta_{1,t}$ in model (3.3) is 10.973% with $t = 3.771$, consistent with previous results. The effect of return gap is significant, but the effect of TR^2 is negative and significant, -0.122% with $t = -3.123$. That is, greater selectivity indicates higher inflows. Similar results are obtained when we consider models (3.4) and (3.5).

Overall, all these results demonstrate the significant cross-fund predictability of the holding-linked network for fund performance and flows.

3.5 Persistent Effect of Holding-Linked Network

We next examine the persistence of the cross-fund predictability of the holding-linked network for fund performance and flows.

Specifically, we consider the cross-fund predictability of the holding-linked network at previous quarters on fund's current month performance and flows. More specifically, we estimate the model:

$$\mathrm{CS}_t^i = \alpha_t + \beta_{0,t} CS_{t-1}^i + \beta_{1,t}^{t-1} \frac{1}{\sum_{j=1}^{N_{t-1})} I(s_{t-1}^{i,j} > 0)} \sum_{j=1}^{N_{t-1}} s_{t-1}^{i,j} CS_{t-1}^j$$
$$+ \sum_{l=2}^{5} \beta_{1,t}^{t-3(l-1)} \frac{1}{\sum_{j=1}^{N_{t-3(l-1)}} I(s_{t-3(l-1)}^{i,j} > 0)} \sum_{j=1}^{N_{t-3(l-1)}} s_{t-3(l-1)}^{i,j} CS_{t-3(l-1)}^j + X_{i,t-1}^{'} \gamma_t + \epsilon_{i,t}, \tag{3.6}$$

where $\beta_{1,t}^{t-l}$, $l = 1, 3, 6, 9, 12$ describe cross-sectional predictability of fund i's linked neighbors at the previous 1, 3, 6, 9, and 12 months. Similar models as (3.6) are estimated with the dependent variable as CT_t^i and $Flow_t^i$. The controlled fund and manager's characteristics $X_{i,t}$ are the same as used in Section 3.4.

Model (3.6) is estimated using the same method as in previous section, that is, following [78] with a monthly cross-sectional regression and Newey-West adjusted standard errors using the optimal lag. In the estimation, the lagged network terms are orthogonalized to control the possible correlations between the lagged network terms.

[Table 3.7 inserted here]

The estimation results are presented in columns (1) to (3) of Table 3.7. When CS is the dependent variable, the cross-fund predictability of the holding-linked funds is significant at the previous one-month and three-month. Specifically, the mean coefficients $\beta_{1,t}^{t-1}$ is 5.536% with $t = 5.582$, consistent with previous results. The mean coefficients $\beta_{1,t}^{t-3}$ is negative and significant, -9.338% with $t = -5.248$. The magnitude of $\beta_{1,t}^{t-3}$ is not smaller than $\beta_{1,t}^{t-1}$. That is, lower performance CS of linked funds in the previous quarter indicates higher performance at current month. This result implies that managers have to adjust their portfolios at least quarterly to improve performance, but not too frequently such as one month.

When CT is the dependent variable, $\beta_{1,t}^{t-1}$ is not significant, same as the results of Section 3.4, however, $\beta_{1,t}^{t-3}$ is positive and significant, 3.758% with $t = 2.833$. That is, the holding-linked network has delayed impact in the prediction of CT. Higher performance CT of linked funds in the previous quarter indicates higher performance at the current month.

When fund flows are the dependent variable, only $\beta_{1,t}^{t-1}$ is positive and significant. That is, only the previous month holding-linked funds have significant cross-fund predictability for fund flows.

In addition to model (3.6), we take the lagged impact of all non-linked funds and consider the following model:

$$\begin{aligned}
\mathrm{CS}_t^i &= \alpha_t + \beta_{0,t} CS_{t-1}^i + \beta_{1,t}^{t-1} \frac{1}{\sum_{j=1}^{N_{t-1)}} I(s_{t-1}^{i,j} > 0)} \sum_{j=1}^{N_{t-1}} s_{t-1}^{i,j} CS_{t-1}^j \\
&+ \sum_{l=2}^{5} \beta_{1,t}^{t-3(l-1)} \frac{1}{\sum_{j=1}^{N_{t-3(l-1)}} I(s_{t-3(l-1)}^{i,j} > 0)} \sum_{j=1}^{N_{t-3(l-1)}} s_{t-3(l-1)}^{i,j} CS_{t-3(l-1)}^j \\
&+ \beta_{2,t}^{t-1} \frac{1}{\sum_{j=1}^{N_{t-1}} I(s_{t-1}^{i,j} = 0)} \sum_{j=1}^{N_{t-1}} I(s_{t-1}^{i,j} = 0) CS_{t-1}^j \\
&+ \sum_{l=2}^{5} \beta_{2,t}^{t-3(l-1)} \frac{1}{\sum_{j=1}^{N_{t-3(l-1)}} I(s_{t-3(l-1)}^{i,j} = 0)} \sum_{j=1}^{N_{t-3(l-1)}} I(s_{t-3(l-1)}^{i,j} = 0) CS_{t-3(l-1)}^j + X_{i,t-1}^{'} \gamma_t + \epsilon_{i,t},
\end{aligned} \tag{3.7}$$

where $\beta_{1,t}^{t-l}$ and $\beta_{2,t}^{t-l}$ represent the effect of holding linked and non-linked funds at previous l-month in the cross-fund prediction of CS, respectively. Similar models as (3.7) are estimated with the dependent variable as CT_t^i and fund flows. The estimation results are summarized in columns (4) to (6) of Table 3.7.

The linked funds' effect $\beta_{1,t}^{t-l}$ in prediction of fund performance and flows are similar as the results from model (3.6). For the cross-fund predictability of the non-linked funds, when CS is the dependent variable, $\beta_{2,t}^{t-1}$ is negative and significant, -0.119% (t=-4.629), while $\beta_{2,t}^{t-3}$ is positive and significant 0.274% (t=5.262), consistent with our previous results. When CT is the dependent variable, $\beta_{2,t}^{t-6}$ is -0.136% with $t = -2.060$, indicating the negative and significant impact from previous half year non-linked funds and the delayed effect of the holding-linked network in the cross-fund predictability of CT. For fund flows, again, only $\beta_{1,t}^{t-1}$ is significant.

Furthermore, similar as model (3.5), we consider the model

$$\text{CS}_t^i = \alpha_t + \beta_{0,t} CS_{t-1}^i + \beta_{3,t}^{t-1} \frac{1}{\sum_{j=1}^{N_{t-1})} I(s_{t-1}^{i,j} \neq 1)} \sum_{j=1}^{N_{t-1}} (1 - s_{t-1}^{i,j}) CS_{t-1}^j$$
$$+ \sum_{l=2}^{5} \beta_{3,t}^{t-3(l-1)} \frac{1}{\sum_{j=1}^{N_{t-3(l-1)}} I(s_{t-3(l-1)}^{i,j} \neq 1)} \sum_{j=1}^{N_{t-3(l-1)}} (1 - s_{t-3(l-1)}^{i,j}) CS_{t-3(l-1)}^j + X_{i,t-1}^{'} \gamma_t + \epsilon_{i,t}. \tag{3.8}$$

The estimation results are summarized in columns (7) to (9) of Table 3.7. For CS and CT, similar results are obtained as model (3.6). It is worth noting that the cross-fund predictability of the previous dissimilarity network for fund flows is significant until previous 12 months, although $\beta_{3,t}^{t-1}$ becomes insignificant in the presence of previous quarters' network effects.

To summarize, the holding-linked network has significant and persistent predictability for fund performance and flows, especially for fund flows.

3.6 What Relate to Fund Holding Similarity?

In this section, we examine the fund and manager's characteristics that relate to the similarity in holdings between funds by regressing $Similarity_t^i$ defined in equation (3.2) on the lagged fund and manager's characteristics that are used in model (3.3). The effects of fund characteristics on $Similarity_t^i$ are estimated by a panel regression, with the standard errors clustered by time and fund.

The results, presented in Table 3.8, show that a fund's similarity in holdings with others is related to some characteristics. The coefficient of flow is negative and significant, suggesting that funds with high similarity in holdings with others is associated with low flow, consistent with results in Panel B of Table 3.2. Similarity is an increasing and concave function of fund size (in logarithm), as evident from the positive and negative coefficients of $\log(\text{TNA})$ and of $\log^2(\text{TNA})$, respectively. The coefficient of expense ratio is negative and significant, suggesting that lower similarity is associated with higher expenses. This may be so because atypical holdings incurs higher cost (e.g., in the acquisition and analysis of information) and because investors may be willing to pay more for funds when it is harder for them to replicate their strategies. The coefficient of fund age is positive and significant, suggesting that old funds tend to make similar portfolio decisions. The coefficient of manager tenure is negative and significant, suggesting that younger managers tend to herd when selecting their portfolios ([58]). The coefficient of number of held stocks is positive and significant, suggesting that more stocks chosen in the portfolios increase the similarity with others. Fund turnover is insignificant, suggesting that funds' similarity in holdings is not associated with trading frequency.

[Table 3.8 inserted here]

As a robustness check, we re-estimate the model using the Fama-MacBech method. The results indicate that the coefficients of $\log(\text{TNA})$, $\log^2(\text{TNA})$, expense ratio, age, manager tenure, and number of held stocks retain their sign and significance.

3.7 Conclusion

We propose a measurement of similarity between funds' portfolios based on Euclidean geometry, resulting a holding-linked network of funds. We find that funds with high similarity in holdings with others tend to have low return before or after adjusting the fees. Sorting funds by their similarity with other funds and their $alpha$s into twenty-five (5×5) fund portfolios, we find that the highest $similarity$ and lowest $alpha$ portfolio minus the lowest $similarity$ and lowest $alpha$ portfolio produces an annual $alpha$ of 4.458% with $t = 4.926$, indicating the significant predictability of funds' similarities for fund performance.

To further examine the predictability of funds' similarities for fund performance and flows, we propose a network based regression by taking the holding-linked network structure into consideration. We find that the holding-linked network has significant and persistent impact in the cross-fund predictability for fund performance and flows. The previous one-month and three-month holding-linked networks are significant in the cross-fund prediction of fund performance measured by CS, and the previous three-month and six-month network are significant in the prediction of fund performance measured by CT. For fund flows, the previous one-month to one-year holding-linked networks demonstrate significant effect in the cross-fund prediction.

All together, this study proposes a new and convenient way of predicting mutual fund performance and flows through the holding-linked network of funds.

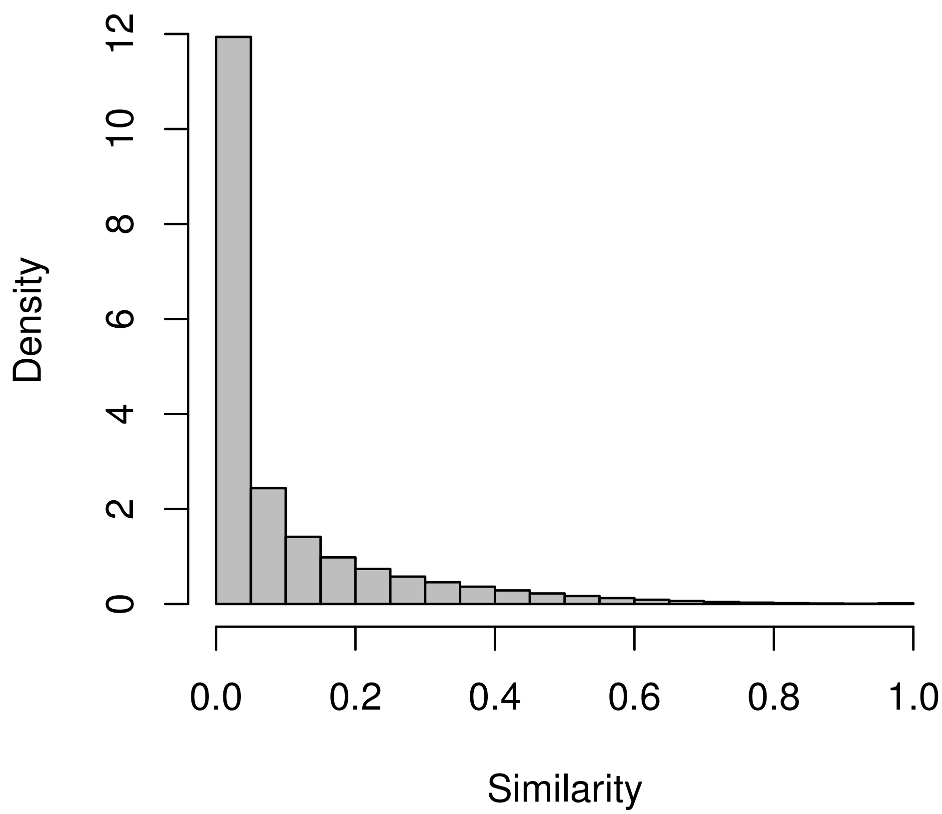

Figure 3.1: Distribution of similarity between funds.
This figure presents the distribution of similarity between funds. Similarity is computed as $Similarity_t^{i,j} = (H_{i,t}^{i,j} \cdot H_{j,t}^{i,j})/(\|H_{i,t}^{i,j}\|_2 \|H_{j,t}^{i,j}\|)$. Details are provided in Subsection 3.2.2 of Section 3.2.

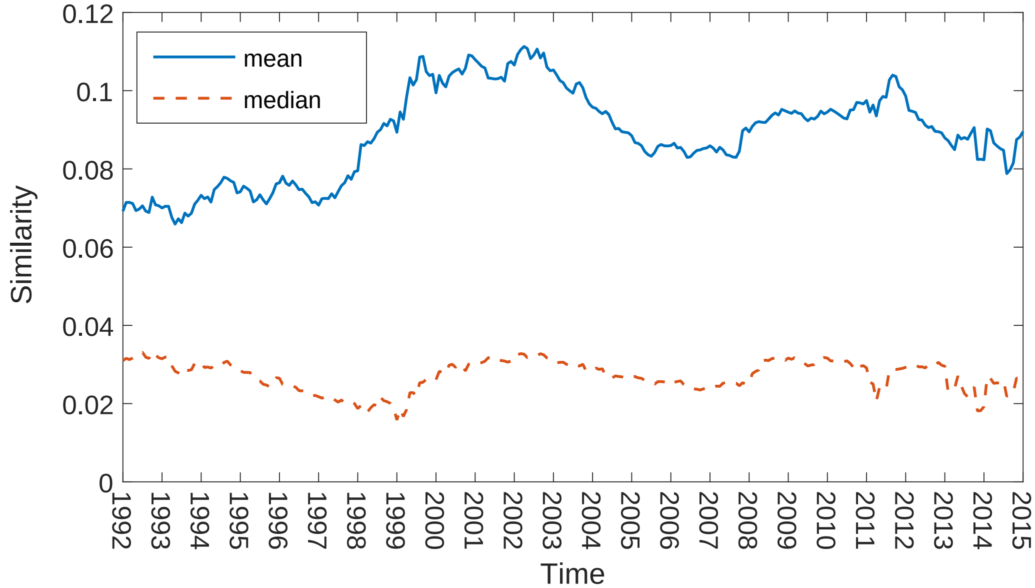

Figure 3.2: Cross-sectional mean and median similarity between funds.
This figure presents the time series of cross-sectional mean and median similarity between funds over the whole sample period from December 1992 to December 2015. The solid line shows the mean similarity and the dashed line shows the median similarity.

Table 3.1: Summary Statistics.

This table reports summary statistics for our sample funds over the whole sample period from December 1992 to December 2015. Each fund's total net asset (TNA), age, expense ratio, turnover ratio, manager tenure, and flow are obtained from CRSP. The fund flow equals the dollar inflows or outflows over the month scaled by the fund's previous month total net assets. All these variables are winsorized by the top and bottom 0.5% tails.

	Mean	Median	Minimum	Maximum
	Mean	Median	Min	Max
Number of unique funds	1,803.740	1,919	602.000	2,569
Number of held stocks	143.083	77.000	11.000	1,940
TNA($millions)	904.159	201.627	5.000	21,944.410
Fund age(years)	11.151	8.467	0.584	64.993
Expense ratio(%)	1.200	1.180	0.007	3.267
Turnover(%)	80.843	61.288	2.010	615.000
Manager Tenure(years)	6.058	4.750	0.334	30.235
Fund flow (%)	0.461	-0.119	-29.847	78.890

Table 3.2: Holding Similarity.

Panel A reports the summary statistics of the similarity between funds over the whole sample period (December 1992 – December 2015) and two evenly divided sub-sample periods (December 1992 – December 2003 and January 2004 – December 2015), including mean, median, minimum, maximum, and standard deviation (STD). Panel B reports the average fund characteristics in each portfolio from sorting all sample funds into quintiles every month over the whole sample period. From the lowest similarity portfolio (Low S) to the highest similarity portfolio (High S), the average fund net return, gross return, number of held stocks, total net assets (TNA), age, expense ratio, turnover, manager tenure, and flow in each portfolio are presented.

Panel A: Summary Statistics of Similarity

Sample Period	Mean	Median	Minimum	Maximum	STD
1992/12 – 2015/12	0.096	0.030	0	1	0.147
1992/12 – 2003/12	0.099	0.029	0	1	0.153
2004/01 – 2015/12	0.095	0.031	0	1	0.145

Panel B: Fund Characteristics Summary through Similarity Sorting

	Low S	2	3	4	High S
Net Return(%)	1.001	1.028	1.012	0.951	0.968
Gross Return (%)	1.114	1.128	1.105	1.041	1.045
Number of Held Stocks	77.095	120.599	136.376	81.736	209.327
TNA (millions)	246.225	469.313	629.959	757.857	1,046.963
Fund Age (years)	10.062	11.812	14.684	16.034	18.603
Expense Ratio (%)	1.369	1.239	1.142	1.108	0.945
Turnover (%)	77.949	94.360	81.968	76.505	70.861
Manager Tenure (years)	6.441	6.301	6.673	6.375	5.724
Fund Flow (%)	1.018	0.902	0.795	0.556	0.474

Table 3.3: Fund portfolio alpha, based on sorting on lagged similarity and alpha. The table presents the portfolio alpha, annualized, using monthly returns. Portfolios are formed by sorting all funds every month into quintiles by similarity and then by alpha. Portfolio excess returns are regressed on the returns of FFC six factors. For each portfolio, we present alpha, the intercept from the regression and its t-statistics, using robust standard errors ([77]). The sample period of the test months is from 5/1980 to 12/2015. "***","**",and "*" denote significance at the 1%, 5%, and 10%, respectively.

Panel A: Results using net returns

	$Similarity^i_{t-1}$					
$Alpha^i_{t-1}$	Low	2	3	4	High	High-Low
Low	−4.597***	−2.415**	−1.788*	−1.667*	−0.139	4.458***
	(−4.398)	(−2.510)	(−1.946)	(−1.811)	(−0.135)	(4.926)
2	−2.964***	−1.666*	−1.967*	−1.216	−0.396	2.568***
	(−2.666)	(−1.872)	(−1.934)	(−1.176)	(−0.311)	(2.742)
3	−1.846**	−1.956**	−1.296	−1.629*	−0.550	1.296
	(−2.140)	(−2.473)	(−1.610)	(−1.875)	(−0.694)	(1.537)
4	−2.133***	−1.826***	−1.545***	−1.031**	−0.319	1.815**
	(−3.509)	(−3.362)	(−3.415)	(−2.415)	(−0.579)	(2.472)
High	−1.766***	−1.528***	−0.995***	−0.509*	−0.823*	0.943**
	(−4.670)	(−5.058)	(−3.402)	(−1.880)	(−1.886)	(2.029)
All	−1.801**	−1.569**	−1.404**	−1.413***	−1.084***	0.718
	(−2.328)	(−1.997)	(−2.247)	(−4.016)	(−4.013)	(0.897)
High-Low	2.831***	0.887	0.793	1.158	−0.684	
	(2.784)	(0.928)	(0.789)	(1.208)	(−0.642)	

Panel B: Results using gross returns

	$Similarity^i_{t-1}$					
$Alpha^i_{t-1}$	Low	2	3	4	High	High-Low
Low	−2.936***	−1.030	−0.490	−0.369	1.261	4.196***
	(−2.832)	(−1.069)	(−0.536)	(−0.396)	(1.227)	(4.685)
2	−1.569	−0.426	−0.747	−0.003	0.851	2.419***
	(−1.412)	(−0.479)	(−0.734)	(−0.002)	(0.672)	(2.593)
3	−0.524	−0.797	−0.182	−0.549	0.585	1.109
	(−0.609)	(−1.009)	(−0.230)	(−0.634)	(0.739)	(1.318)
4	−0.870	−0.665	−0.445	0.043	0.769	1.639**
	(−1.431)	(−1.226)	(−0.986)	(0.100)	(1.398)	(2.233)
High	−0.622*	−0.474	−0.075	0.310	0.093	0.715
	(−1.654)	(−1.570)	(−0.256)	(1.138)	(0.212)	(1.542)
All	−0.409	−0.321	−0.245	−0.279	−0.132	0.277
	(−0.529)	(−0.408)	(−0.393)	(−0.793)	(−0.488)	(0.347)
High-Low	2.313**	0.556	0.415	0.678	−1.168	
	(2.287)	(0.581)	(0.414)	(0.700)	(−1.096)	

Table 3.4: The impact of similarity on Fund Performance.

Estimation results of the benchmark model and models (3.3), (3.4), and (3.5). The dependent variables are monthly measures of two fund performance measures proposed by [61]: CS, characteristic selectivity, and CT, characteristic timing. Similarity represents the effect of holding-linked funds, i.e., $\beta_{1,t}$ in models (3.3) and (3.4). Complete dissimilarity represents the effect of holding non-linked funds, i.e., $\beta_{2,t}$ in models (3.3) and (3.4). Dissimilarity represents the effect of dissimilarity network, i.e., $\beta_{3,t}$ in model (3.5). The variable, number of held stocks, is scaled in hundred in regression. Other explanatory variables are as described in Table 3.2, and we include six dummy variables in regression. The cross-sectional estimation is by the Fama-MacBech method over the 276 monthly test periods, 1993–2015, with Newey-West standard errors using optimal lag. The means of the coefficients and their t-statistics (in parenthesis) are presented. "***","**",and "*" denote significance at the 1%, 5%, and 10%, respectively.

	Dependent variable							
	CS				CT			
Explanatory variables, lagged	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Similarity ($\beta_{1,t}$)		4.672***	4.063***			1.000	0.922	
		(5.001)	(4.700)			(1.238)	(1.176)	
Complete Dissimilarity ($\beta_{2,t}$)			−0.156***				0.004	
			(−3.481)				(0.067)	
Dissimilarity ($\beta_{3,t}$)				−5.316***				−1.326
				(−4.796)				(−1.295)
Lag Performance ($\beta_{0,t}$)	0.693***	0.678***	0.676***	0.679***	0.029*	0.028*	0.028*	0.028*
	(59.404)	(62.654)	(63.179)	(61.998)	(1.679)	(1.695)	(1.758)	(1.704)
Flow	−0.006	−0.005	−0.003	−0.005	0.005	0.001	0.001	0.001
	(−0.658)	(−0.575)	(−0.479)	(−0.624)	(0.709)	(0.111)	(0.185)	(0.169)
log(TNA)	−0.053	−0.078	−0.078*	−0.079*	−0.013	−0.045	−0.058	−0.042
	(−1.059)	(−1.646)	(−1.658)	(−1.651)	(−0.294)	(−1.078)	(−1.318)	(−1.001)
$\log^2$(TNA)	0.002	0.004	0.004	0.004	0.001	0.003	0.004	0.003
	(0.525)	(0.897)	(0.934)	(0.889)	(0.252)	(0.756)	(0.997)	(0.699)
Expense Ratio	−0.028	−0.009	0.004	−0.006	0.037	0.012	−0.017	0.022
	(−0.291)	(−0.103)	(0.051)	(−0.069)	(0.784)	(0.205)	(−0.310)	(0.381)
Turnover	0.000	0.000	0.000	0.000	0.002**	0.002**	0.002**	0.002**
	(−0.004)	(−0.377)	(−0.261)	(−0.334)	(2.273)	(2.305)	(2.282)	(2.353)
log(Age)	0.060*	0.055**	0.060**	0.053*	0.030	0.040	0.027	0.039
	(1.813)	(1.992)	(2.191)	(1.919)	(0.673)	(0.943)	(0.612)	(0.913)
log(Mgr Tenure)	−0.016	0.001	−0.002	0.001	0.028	0.050	0.059	0.049
	(−0.448)	(0.043)	(−0.077)	(0.046)	(0.512)	(0.797)	(0.940)	(0.769)
Number of held stocks	0.004	−0.005	−0.008	−0.009	−0.001	−0.007	−0.011	−0.009
	(−0.496)	(−0.739)	(−1.011)	(−1.010)	(−0.049)	(−0.545)	(−0.750)	(−0.700)
Regression R^2	0.613	0.622	0.625	0.621	0.189	0.210	0.216	0.209

Table 3.5: The impact of similarity on Fund Flow.
Estimation results of the benchmark model and models (3.3), (3.4), and (3.5) with the dependent variable as fund flow. Similarity represents the effect of holding-linked funds, i.e., $\beta_{1,t}$ in models (3.3) and (3.4). Complete dissimilarity represents the effect of holding non-linked funds, i.e., $\beta_{2,t}$ in models (3.3) and (3.4). Dissimilarity represents the effect of dissimilarity network, i.e., $\beta_{3,t}$ in model (3.5). The variable, number of held stocks, is scaled in hundred in regression. Other explanatory variables are as described in Table 3.2, and we include six dummy variables in regression. The cross-sectional estimation is by the Fama-MacBech method over the 276 monthly test periods, 1993–2015, with Newey-West standard errors using optimal lag. The means of the coefficients and their t-statistics (in parenthesis) are presented. "***","**",and "*" denote significance at the 1%, 5%, and 10%, respectively.

Explanatory variables, lagged	Dependent variable: $Flow_t^i$			
Similarity ($\beta_{1,t}$)		8.252^{***}	8.200^{***}	
		(2.939)	(2.955)	
Complete Dissimilarity ($\beta_{2,t}$)			-0.018	
			(-0.030)	
Dissimilarity ($\beta_{3,t}$)				-13.481^{***}
				(-3.841)
Flow ($\beta_{0,t}$)	-0.054^{***}	-0.056^{***}	-0.057^{***}	-0.064^{***}
	(-8.622)	(-8.314)	(-8.345)	(-8.425)
log(TNA)	-4.08^{***}	-0.446^{***}	-0.446^{***}	-0.447^{***}
	(-7.659)	(-8.161)	(-8.762)	(-8.506)
$\log^2$(TNA)	0.034^{***}	0.037^{***}	0.037^{***}	0.037^{***}
	(7.058)	(7.566)	(7.579)	(7.459)
Expense Ratio	-0.104	-0.131	-0.133	-0.135
	(-1.073)	(-1.390)	(-1.423)	(-1.435)
Turnover	-0.001^{***}	-0.001^{***}	-0.001^{***}	-0.001^{***}
	(-3.067)	(-3.059)	(-3.081)	(-3.002)
log(Age)	-0.879^{***}	-0.877^{***}	-0.873^{***}	-0.875^{***}
	(-11.992)	(-12.453)	(-12.490)	(-12.436)
log(Mgr Tenure)	0.051^{**}	0.047^{*}	0.046^{*}	0.045^{*}
	(1.991)	(1.893)	(1.860)	(1.801)
Net Return	0.121^{***}	0.098^{***}	0.104^{***}	0.101^{***}
	(6.593)	(5.560)	(5.858)	(5.646)
Number of held stocks	0.023^{*}	0.014	0.014	0.017
	(1.745)	(1.319)	(1.170)	(1.467)
Regression R^2	0.044	0.049	0.051	0.049

Table 3.6: Comparison with other fund performance predictors.

Estimation results of models (3.3), (3.4), and (3.5). The dependent variables are monthly measures of two fund performance measures proposed by [61]: CS, characteristic selectivity, and CT, characteristic timing, and fund flow. Similarity represents the effect of holding-linked funds, i.e., $\beta_{1,t}$ in models (3.3) and (3.4). Complete dissimilarity represents the effect of holding non-linked funds, i.e., $\beta_{2,t}$ in models (3.3) and (3.4). Dissimilarity represents the effect of dissimilarity network, i.e., $\beta_{3,t}$ in model (3.5). The explanatory variable "Return Gap" proposed by [59] and the variable TR^2 proposed by [62] are included. The variable, number of held stocks, is scaled in hundred in regression. Other explanatory variables are as described in Table 3.2, and we include six dummy variables in regression. The cross-sectional estimation is by the Fama-MacBech method over the 276 monthly test periods, 1993–2015, with Newey-West standard errors using optimal lag. The means of the coefficients and their t-statistics (in parenthesis) are presented. "***","**",and "*" denote significance at the 1%, 5%, and 10%, respectively.

Explanatory variables	Dependent variable								
	CS			CT			Fund Flow		
Similarity	4.743***	4.179***		1.364	1.354		10.973***	10.989***	
	(5.015)	(4.686)		(1.609)	(1.630)		(3.771)	(3.587)	
Complete Dissimilarity		−0.149***			0.049			0.244	
		(−3.800)			(0.823)			(0.321)	
Dissimilarity			−5.675***			−1.907*			−16.313***
			(−5.008)			(−1.774)			(−4.433)
Lag Dependent Variable	0.668***	0.666***	0.667***	0.023	0.023	0.023	−0.050***	−0.050***	−0.059***
	(57.626)	(58.116)	(57.799)	(1.478)	(1.475)	(1.445)	(−6.663)	(−6.807)	(−7.332)
Return Gap	0.006*	0.006*	0.006*	0.006	0.005	0.005	0.013***	0.013***	0.014***
	(1.768)	(1.676)	(1.770)	(1.460)	(1.354)	(1.399)	(3.618)	(3.707)	(3.708)
TR^2	−0.044	−0.042	−0.045	0.188**	0.185**	0.192**	−0.122***	−0.124***	−0.121***
	(−0.959)	(−0.992)	(−0.983)	(2.340)	(2.464)	(2.585)	(−3.123)	(−3.051)	(−3.071)
Fund characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regression R^2	0.632	0.635	0.631	0.231	0.237	0.231	0.070	0.073	0.069

Table 3.7: Persistence of similarity impact on Fund Performance and Flow.

Estimation results of models (3.6), (3.7), and (3.8). The dependent variables are monthly measures of two fund performance measures proposed by [61]: CS, characteristic selectivity, and CT, characteristic timing, and fund flow. For $l = 1, 3, 6, 9, 12$, $\beta_{1,t}^{t-l}$ represent the effects of holding linked funds in models (3.6) and (3.7), $\beta_{2,t}^{t-l}$ represents the effects of holding non-linked funds in models (3.6) and (3.7), and $\beta_{3,t}^{t-l}$ represents the effects from dissimilarity network in model (3.8). All the lagged variables of linked and non-linked funds' effects, and dissimilarity network effects in each model are orthogonalized to control for the correlation among them in regression estimation. The variable, number of held stocks, is scaled in hundred in regression. Other explanatory variables are as described in Table 3.2, and we include six dummy variables in regression. The cross-sectional estimation is by the Fama-MacBech method over the 276 monthly test periods, 1993–2015, with Newey-West standard errors using optimal lag. The means of the coefficients and their t-statistics (in parenthesis) are presented. "***","**",and "*" denote significance at the 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Similarity			Similarity & Complete Dissimilarity			Dissimilarity		
Explanatory				Dependent variable					
variables	CS	CT	Flow	CS	CT	Flow	CS	CT	Flow
$\beta_{1,t}^{t-1}$	5.536***	1.241	8.946***	5.297***	1.033	9.442***			
	(5.582)	(1.315)	(3.225)	(5.670)	(1.168)	(3.580)			
$\beta_{1,t}^{t-3}$	−9.388***	3.758***	1.838	−8.763***	2.909**	1.128			
	(−5.248)	(2.833)	(1.063)	(−4.770)	(2.407)	(0.651)			
$\beta_{1,t}^{t-6}$	−0.476	1.006	1.228	−0.474	0.767	0.426			
	(−0.672)	(0.840)	(0.549)	(−0.683)	(0.672)	(0.212)			
$\beta_{1,t}^{t-9}$	−0.790	−0.559	−1.394	−0.755	−0.082	−1.063			
	(−0.834)	(−0.827)	(−1.045)	(−0.856)	(−0.123)	(−0.803)			
$\beta_{1,t}^{t-12}$	−0.640	0.089	−0.670	−0.634	−0.288	−1.072			
	(−0.656)	(0.082)	(−0.458)	(−0.673)	(−0.256)	(−0.710)			
$\beta_{2,t}^{t-1}$				−0.119***	−0.032	0.599			
				(−4.629)	(−0.435)	(0.983)			
$\beta_{2,t}^{t-3}$				0.274***	−0.044	0.121			
				(5.262)	(−0.730)	(0.415)			
$\beta_{2,t}^{t-6}$				−0.024	−0.136**	0.113			
				(−0.701)	(−2.060)	(0.303)			
$\beta_{2,t}^{t-9}$				0.008	0.009	−0.302			
				(0.392)	(0.151)	(−0.699)			
$\beta_{2,t}^{t-12}$				−0.024	−0.018	0.052			
				(−0.865)	(−0.319)	(0.144)			
$\beta_{3,t}^{t-1}$							−5.676***	1.109	−9.156
							(−4.323)	(0.479)	(−1.189)
$\beta_{3,t}^{t-3}$							9.520***	−4.112**	−8.807***
							(6.499)	(−2.283)	(−2.991)
$\beta_{3,t}^{t-6}$							−0.074	0.630	−7.496***
							(−0.229)	(0.460)	(−3.167)
$\beta_{3,t}^{t-9}$							0.266	0.109	−4.119**
							(0.743)	(0.068)	(−2.536)
$\beta_{3,t}^{t-12}$							0.187	−1.208	−5.017***
							(0.485)	(−0.941)	(−2.788)
Lag Dependent	0.675***	0.021	−0.058***	0.673***	0.022	−0.058***	0.678***	0.022	−0.063***
Variable	(59.764)	(1.262)	(−8.248)	(60.110)	(1.354)	(−8.117)	(61.724)	(1.259)	(−8.265)
Fund Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regression R^2	0.639	0.255	0.062	0.644	0.270	0.071	0.631	0.243	0.063

Table 3.8: Determinants of fund similarity.

The dependent variable is $Similarity_t^i$, which represents the similarity in holdings between fund i and all other funds at month t. The variable, number of held stocks, is scaled in hundred in regression. Other explanatory variables are as in Table 3.2. The estimation is panel regression, with standard errors clustered by time and funds. The coefficients and their t-statistics (in parenthesis) are presented. "***","**",and "*" denote significance at the 1%, 5%, and 10%, respectively.

Explanatory variables, lagged	Dependent variable: $Similarity_t^i$
Flow	-0.0001^{***}
	(-4.7469)
log(TNA)	0.0061^{***}
	(4.1652)
$\log^2$(TNA)	-0.0003^{*}
	(-1.6971)
Expense ratio	-0.0205^{***}
	(-10.1527)
Turnover	0.0000
	(1.1364)
log(Fund age)	0.0023^{**}
	(2.0011)
log(Mgr Tenure)	-0.0073^{***}
	(-7.2531)
Number of held stocks	0.0041^{***}
	(8.1206)
Observations	248,683
Regression R^2	0.4904

Appendices

APPENDIX A

GLOBALIZATION OF LOCAL FACTORS:IMPLICATIONS FOR THE ASSET PRICING

Table A.1: Datastream active- and dead-firm constituent lists.

Country	Active-firm constituent list	Dead-firm constituent list
Canada	FTORO	DEADCN1
		DEADCN2
		DEADCN3
		DEADCN4
		DEADCN5
		DEADCN6
Japan	FJAP	DEADJP
U.K.	FBRIT	DEADUK

Table A.2: Globalization of local factors, measured by R^2.

For each sample country, each of the six local factors (MKT, SMB, HML, MOM, RMW, CMA) is regressed on the global factors over five-year rolling windows for the whole sample period of July 1990 – December 2017. The time-series average R^2 from the regressions for each local factor is reported. In addition, we formally test the significance of the average R^2 and whether the R^2s change over time. The adjusted t-statistics are reported, which considers the serial correlation of R^2s. “*”, “**”, and “***” indicate significance at the 10%, 5%, and 1% levels, respectively. In Panel A, the global factors in the regressions of the local factors are constructed using the equal-weighted average of the local factors of Australia, Canada, France, Germany, Hong Kong, Japan, Switzerland, the U.K., and the U.S. In Panel B, the global factors in the regressions of the local factors are constructed through pooling all firms from these nine countries.

Panel A: Equal-weighted local factors as global factors							
Country	Local Factor	MKT	SMB	HML	MOM	RMW	CMA
Australia	Mean	0.716	0.354	0.222	0.363	0.408	0.323
	t-stat	14.258***	11.485***	8.238***	6.916***	5.703***	4.289***
	Slope	0.001	0.000	0.000	0.001	0.001	0.001
	t-stat	2.733***	1.019	0.311	7.534***	1.790*	2.110**
Canada	Mean	0.739	0.476	0.342	0.536	0.587	0.357
	t-stat	22.513***	7.239***	8.281***	15.567***	14.433***	7.961***
	Slope	0.001	0.000	0.000	0.001	0.001	0.000
	t-stat	2.152**	0.546	−1.134	4.188***	1.937*	0.030
France	Mean	0.787	0.580	0.396	0.577	0.167	0.252
	t-stat	12.885***	16.344***	6.583***	6.919***	2.208**	5.835***
	Slope	0.001	0.001	0.001	0.001	−0.001	0.001
	t-stat	3.932***	4.236***	2.757***	1.655*	−1.975**	1.673*
Germany	Mean	0.764	0.483	0.266	0.520	0.249	0.288
	t-stat	11.302***	12.269***	4.165***	6.403***	3.742***	12.670***
	Slope	0.002	0.001	0.002	0.002	−0.001	0.000
	t-stat	5.349***	3.571***	4.660***	1.787*	−0.881	−2.470**
Hong Kong	Mean	0.683	0.450	0.233	0.338	0.301	0.163
	t-stat	16.760***	9.433***	3.657***	9.480***	8.154***	3.455***
	Slope	0.001	−0.001	−0.001	0.001	0.000	0.000
	t-stat	4.055***	−1.834*	−2.763***	1.993**	−0.010	1.310
Japan	Mean	0.451	0.151	0.217	0.279	0.209	0.213
	t-stat	12.857***	4.164***	6.048***	5.040***	5.861***	5.123***
	Slope	0.001	0.000	−0.001	0.001	−0.001	0.000
	t-stat	2.001**	−0.483	−1.906*	3.534***	−4.595***	−0.102
Switzerland	Mean	0.680	0.458	0.238	0.545	0.180	0.337
	t-stat	10.532***	12.023***	4.122***	5.799***	4.472***	6.549***
	Slope	0.002	0.000	0.002	0.002	0.000	0.001
	t-stat	6.642***	0.341	10.706***	2.361**	−0.761	1.418
U.K.	Mean	0.805	0.535	0.421	0.613	0.192	0.278
	t-stat	16.537***	17.455***	7.424***	8.408***	4.437***	5.432***
	Slope	0.001	0.000	0.001	0.001	0.000	0.001
	t-stat	4.530***	0.673	1.719*	1.861*	−0.770	0.838
U.S.	Mean	0.763	0.256	0.460	0.609	0.418	0.354
	t-stat	14.158***	2.614***	6.688***	7.490***	4.459***	6.920***
	Slope	0.001	−0.002	0.000	0.002	0.000	0.000
	t-stat	1.761*	−3.797***	−0.353	1.913*	−0.441	0.180

Panel B: All firms from all countries to construct global factors							
Country	Local Factor	MKT	SMB	HML	MOM	RMW	CMA
Australia	Mean	0.629	0.051	0.059	0.227	0.203	0.113
	t-stat	10.418***	1.820*	1.720*	3.530***	5.259***	2.377**
	Slope	0.001	−0.001	0.000	0.001	0.000	0.001
	t-stat	2.808***	−1.776*	1.472	2.707***	0.292	3.283***
Canada	Mean	0.698	0.229	0.152	0.306	0.314	0.165
	t-stat	14.783***	3.227***	3.003***	5.504***	4.960***	4.024***
	Slope	0.001	−0.001	0.000	0.001	0.000	0.000
	t-stat	1.748*	−1.895*	−0.556	2.047**	0.235	0.140
France	Mean	0.760	0.367	0.268	0.475	0.116	0.194
	t-stat	11.128***	10.537***	4.873***	5.866***	1.914*	4.191***
	Slope	0.002	0.001	0.000	0.002	−0.001	0.001
	t-stat	4.308***	1.725*	0.799	2.036**	−1.304	2.143**
Germany	Mean	0.747	0.299	0.133	0.433	0.150	0.108
	t-stat	9.001***	9.944***	3.011***	5.478***	2.292**	2.667***
	Slope	0.002	0.001	0.001	0.002	0.000	−0.001
	t-stat	4.825***	1.924*	1.685*	2.121**	−0.832	−2.741***
Hong Kong	Mean	0.563	0.106	0.047	0.212	0.069	0.046
	t-stat	8.540***	2.733***	2.436**	3.843***	2.212**	1.559
	Slope	0.002	−0.001	0.000	0.002	0.000	0.000
	t-stat	2.932***	−2.457**	−0.374	6.525***	1.817*	0.876
Japan	Mean	0.777	0.203	0.311	0.362	0.324	0.291
	t-stat	18.862***	3.429***	6.136***	7.381***	5.597***	6.999***
	Slope	−0.001	−0.001	0.000	0.000	−0.001	−0.001
	t-stat	−2.847***	−4.725***	−0.713	0.126	−3.871***	−1.274
Switzerland	Mean	0.630	0.286	0.123	0.405	0.038	0.184
	t-stat	8.612***	7.672***	2.644***	3.876***	1.183	3.572***
	Slope	0.002	0.000	0.001	0.002	0.000	0.001
	t-stat	4.584***	0.809	2.605***	2.290**	1.172	2.169**
U.K.	Mean	0.785	0.373	0.304	0.548	0.119	0.137
	t-stat	17.481***	12.785***	5.065***	6.806***	3.196***	5.541***
	Slope	0.001	0.000	0.001	0.002	0.000	0.000
	t-stat	4.083***	0.550	1.121	2.482**	−1.865*	0.645
U.S.	Mean	0.941	0.772	0.642	0.775	0.690	0.461
	t-stat	68.506***	20.594***	12.866***	9.578***	8.370***	9.822***
	Slope	0.000	0.000	−0.001	0.002	0.001	−0.001
	t-stat	1.754*	−1.412	−1.641	2.161**	1.441	−4.864***

Table A.3: Variance decomposition of size and R^2 sorted portfolio returns.

For Canada, Japan, the U.K., and the U.S., 25 size and R^2 sorted portfolios' excess returns are regressed on the global and pure local factors. The proportions of the portfolio return variance explained by the global factors, pure local factors, and the residual component in the regressions are reported.

		Size-R^2 sorted portfolios																									
		Small					2					3					4					Big					
		Seg	2	3	4	Int	Seg	2	3	4	Int	Seg	2	3	4	Int	Seg	2	3	4	Int	Seg	2	3	4	Int	
Country	Component	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	Average
Canada	Global	0.34	0.33	0.34	0.32	0.23	0.35	0.46	0.40	0.38	0.40	0.51	0.50	0.55	0.54	0.54	0.46	0.54	0.56	0.59	0.67	0.20	0.33	0.32	0.55	0.73	0.44
	Pure Local	0.21	0.21	0.25	0.15	0.15	0.23	0.24	0.24	0.25	0.14	0.13	0.19	0.22	0.17	0.16	0.16	0.13	0.20	0.20	0.14	0.16	0.17	0.21	0.14	0.17	0.18
	Idiosyncratic	0.45	0.47	0.41	0.53	0.63	0.43	0.30	0.37	0.37	0.47	0.36	0.31	0.23	0.29	0.30	0.38	0.33	0.23	0.21	0.19	0.65	0.50	0.48	0.31	0.10	0.37
Japan	Global	0.23	0.25	0.24	0.23	0.22	0.22	0.27	0.28	0.26	0.24	0.23	0.26	0.26	0.26	0.28	0.19	0.21	0.26	0.29	0.33	0.23	0.29	0.34	0.44	0.53	0.27
	Pure Local	0.64	0.63	0.65	0.66	0.64	0.65	0.64	0.62	0.65	0.65	0.65	0.64	0.64	0.63	0.60	0.63	0.64	0.59	0.59	0.53	0.49	0.48	0.45	0.44	0.36	0.59
	Idiosyncratic	0.13	0.12	0.11	0.11	0.14	0.13	0.08	0.09	0.09	0.10	0.12	0.10	0.10	0.11	0.12	0.18	0.15	0.14	0.11	0.14	0.28	0.23	0.21	0.13	0.11	0.13
U.K.	Global	0.44	0.39	0.36	0.28	0.13	0.44	0.48	0.49	0.44	0.43	0.45	0.53	0.55	0.57	0.59	0.36	0.46	0.62	0.65	0.71	0.14	0.37	0.54	0.66	0.79	0.47
	Pure Local	0.29	0.35	0.28	0.17	0.15	0.35	0.34	0.33	0.32	0.28	0.29	0.27	0.30	0.26	0.18	0.17	0.26	0.20	0.20	0.15	0.08	0.17	0.23	0.18	0.15	0.24
	Idiosyncratic	0.27	0.26	0.36	0.54	0.72	0.21	0.18	0.18	0.24	0.29	0.26	0.20	0.15	0.16	0.23	0.47	0.28	0.18	0.15	0.15	0.78	0.46	0.23	0.15	0.06	0.29
U.S.	Global	0.63	0.62	0.64	0.62	0.59	0.71	0.77	0.77	0.76	0.77	0.70	0.75	0.79	0.82	0.83	0.72	0.76	0.79	0.84	0.85	0.51	0.66	0.75	0.83	0.90	0.74
	Pure Local	0.03	0.04	0.03	0.05	0.04	0.05	0.05	0.07	0.08	0.08	0.12	0.13	0.11	0.12	0.09	0.11	0.12	0.13	0.11	0.09	0.05	0.12	0.08	0.07	0.06	0.08
	Idiosyncratic	0.34	0.34	0.33	0.33	0.37	0.24	0.18	0.16	0.16	0.16	0.17	0.12	0.10	0.06	0.08	0.16	0.12	0.08	0.05	0.06	0.44	0.22	0.17	0.09	0.03	0.18

APPENDIX B

DYNAMIC FIRM-SPECIFIC NETWORK AND RISK DIVERSIFICATION

A. Bayesian nonparametric regression To implement the estimation of the Bayesian nonparametric model described in equations (4a), (4b), and (4c), I follow the exact Bayesian estimation steps describes in Section 2.3 of [38] but without the latent process. There are several reasons that I do not consider the latent process as in [38]. First, the latent process is not identifiable. In order to ensure the identification of the latent process, other assumptions are needed. Second, if I add the latent process when the number of firms is very large, the computation is too slow to be implemented by using a regular computer. Third, it is hard to find the economical meaning of the latent process.

To flexibly model the effect of the linkages X_t^κ on β_t^κ, I consider independent Gaussian process priors and set the priors exactly the same as [38]. Then I update the time-varying effects of linkages, β_t^κ, simultaneously for all t by

$$\beta^\kappa \sim N_N(\Sigma_{\beta^\kappa} \begin{bmatrix} \sum_{i=2}^{p} \sum_{j=1}^{i-1} X_{ij\kappa,t_1}(\Omega_{ij,t_1} - 1/2 - \omega_{ig,t_1}\nu_{ij\kappa,t_1}) \\ \vdots \\ \sum_{i=2}^{p} \sum_{j=1}^{i-1} X_{ij\kappa,t_N}(\Omega_{ij,t_N} - 1/2 - \omega_{ig,t_N}\nu_{ij\kappa,t_N}) \end{bmatrix}, \Sigma_{\beta_\kappa})$$

where $\nu_{ij\kappa,t} = \mu_t + X_{ij(-\kappa)}^T \beta_t^{(-\kappa)}$, $t = t_1, \cdots, t_N$, the posterior covariance matrix $\Sigma_{\beta^\kappa} = \{diag(\sum_{i=2}^{p} \sum_{j=1}^{i-1} X_{ij\kappa,t_1}^2 \omega_{ij,t_1}, \cdots, \sum_{i=2}^{p} \sum_{j=1}^{i-1} X_{ij\kappa,t_N}^2 \omega_{ij,t_N}) + K_p^{-1}\}^{-1}$, and the Gaussian process covariance matrix K_p with entries $(K_p)_{ij} = \exp(-\kappa_p \|t_i - t_j\|_2^2)$. The detailed steps in the implementation of the β_t^κ estimation are exactly the same as [38].

REFERENCES

[1] G. Bekaert and C. R. Harvey, "Time-varying world market integration," *The Journal of Finance*, vol. 50, no. 2, pp. 403–444, 1995.

[2] K. Pukthuanthong and R. Roll, "Global market integration: An alternative measure and its application," *Journal of Financial Economics*, vol. 94, no. 2, pp. 214–232, 2009.

[3] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, vol. 33, no. 1, pp. 3–56, 1993.

[4] J. M. Griffin, "Are the fama and french factors global or country specific?" *Review of Financial Studies*, vol. 15, no. 3, pp. 783–803, 2002.

[5] K. Hou, G. A. Karolyi, and B.-C. Kho, "What factors drive global stock returns?" *Review of Financial Studies*, vol. 24, no. 8, pp. 2527–2574, 2011.

[6] E. F. Fama and K. R. French, "Size, value, and momentum in international stock returns," *Journal of Financial Economics*, vol. 105, no. 3, pp. 457–472, 2012.

[7] ——, "A five-factor asset pricing model," *Journal of Financial Economics*, vol. 116, no. 1, pp. 1–22, 2015.

[8] M. M. Carhart, "On persistence in mutual fund performance," *The Journal of Finance*, vol. 52, no. 1, pp. 57–82, 1997.

[9] O. S. Ince and R. B. Porter, "Individual equity return data from thomson datastream: Handle with care!" *Journal of Financial Research*, vol. 29, no. 4, pp. 463–479, 2006.

[10] J. M. Griffin, P. J. Kelly, and F. Nardari, "Do market efficiency measures yield correct inferences? a comparison of developed and emerging markets," *Review of Financial Studies*, vol. 23, pp. 3225–3277, 2010.

[11] I. Chaieb, H. Langlois, O. Scaillet, *et al.*, "Time-varying risk premia in large international equity markets," *Swiss Finance Institute Research Paper*, 2018.

[12] H. Bessembinder, T.-F. Chen, G. Choi, and K. Wei, "Do global stocks outperform us treasury bills?" *Te-Feng and Choi, Goeun and Wei, Kuo-Chiang (John), Do Global Stocks Outperform US Treasury Bills*, 2019.

[13] K.-H. Lee, "The world price of liquidity risk," *Journal of Financial Economics*, vol. 99, no. 1, pp. 136–161, 2011.

[14] T. Shumway, “The delisting bias in crsp data,” *The Journal of Finance*, vol. 52, no. 1, pp. 327–340, 1997.

[15] C. S. Eun, S. Lai, F. A. d. Roon, and Z. Zhang, “International diversification with factor funds,” *Management Science*, vol. 56, no. 9, pp. 1500–1518, 2010.

[16] G. Bekaert, M. Ehrmann, M. Fratzscher, and A. Mehl, “The global crisis and equity market contagion,” *The Journal of Finance*, vol. 69, no. 6, pp. 2597–2649, 2014.

[17] E. F. Fama and K. R. French, “The cross-section of expected stock returns,” *The Journal of Finance*, vol. 47, no. 2, pp. 427–465, 1992.

[18] ——, “Multifactor explanations of asset pricing anomalies,” *The Journal of Finance*, vol. 51, no. 1, pp. 55–84, 1996.

[19] K. Daniel and S. Titman, “Market reactions to tangible and intangible information,” *The Journal of Finance*, vol. 61, no. 4, pp. 1605–1643, 2006.

[20] P. A. Gompers and A. Metrick, “Institutional investors and equity prices,” *The Quarterly Journal of Economics*, vol. 116, no. 1, pp. 229–259, 2001.

[21] W. K. Newey and K. D. West, “A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix,” *Econometrica*, vol. 55, no. 3, pp. 703–708, 1987.

[22] ——, “Automatic lag selection in covariance matrix estimation,” *The Review of Economic Studies*, vol. 61, no. 4, pp. 631–653, 1994.

[23] F. Carrieri, V. Errunza, and K. Hogan, “Characterizing world market integration through time,” *Journal of Financial and Quantitative Analysis*, vol. 42, no. 04, pp. 915–940, 2007.

[24] D. Berger, K. Pukthuanthong, and J. J. Yang, “International diversification with frontier markets,” *Journal of Financial Economics*, vol. 101, no. 1, pp. 227–242, 2011.

[25] J. M. Griffin, X. Ji, and J. S. Martin, “Momentum investing and business cycle risk: Evidence from pole to pole,” *The Journal of Finance*, vol. 58, no. 6, pp. 2515–2547, 2003.

[26] M. R. Gibbons, S. A. Ross, and J. Shanken, “A test of the efficiency of a given portfolio,” *Econometrica: Journal of the Econometric Society*, pp. 1121–1152, 1989.

[27] M. E. H. Arouri, D. K. Nguyen, and K. Pukthuanthong, “An international capm for partially integrated markets: Theory and empirical evidence,” *Journal of Banking and Finance*, vol. 36, no. 9, pp. 2473–2493, 2012.

[28] E. F. Fama and K. R. French, "Value versus growth: The international evidence," *The Journal of Finance*, vol. 53, no. 6, pp. 1975–1999, 1998.

[29] J.-K. Kang and R. Stulz, "Why is there a home bias? an analysis of foreign portfolio equity ownership in japan," *Journal of Financial Economics*, vol. 46, no. 1, pp. 3–28, 1997.

[30] F. Carrieri, I. Chaieb, and V. Errunza, "Do implicit barriers matter for globalization?" *Review of Financial Studies*, vol. 26, no. 7, pp. 1694–1739, 2013.

[31] L. Cohen and A. Frazzini, "Economic links and predictable returns," *The Journal of Finance*, vol. 63, no. 4, pp. 1977–2011, 2008.

[32] M. K. Brunnermeier and L. H. Pedersen, "Market liquidity and funding liquidity," *Review of Financial Studies*, vol. 22, no. 6, pp. 2201–2238, 2009.

[33] S. R. W. G. Geneva Association, "Systemic risk in insurance: Analysis of insurance and financial stability," *Geneva, Switzerland*, vol. 2010, 2010.

[34] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon, "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *Journal of Financial Economics*, vol. 104, no. 3, pp. 535–559, 2012.

[35] D. Rapach, J. Strauss, J. Tu, and G. Zhou, "Industry interdependencies and cross-industry return predictability," *Working Paper*, vol. 2015, 2015.

[36] S. Basu, S. Das, G. Michailidis, and A. K. Purnanandam, "A system-wide approach to measure connectivity in the financial sector," *Working Paper*, vol. 2016, 2016.

[37] N. Meinshausen and P. Buhlmann, "High-dimensional graphs and variable selection with the lasso," *Annals of Statistics*, vol. 34, no. 3, pp. 1436–1462, 2006.

[38] D. Durante and D. B. Dunson, "Nonparametric bayes dynamic modeling of relational data," *Biometrika*, vol. 101, no. 4, pp. 883–898, 2014.

[39] K. Baba, R. Shibata, and M. Sibuya, "Partial correlation and conditional correlation as measures of conditional independence," *Australian and New Zealand Journal of Statistics*, vol. 46, no. 4, pp. 657–664, 2004.

[40] R. Jagannathan and Z. Wang, "The conditional capm and the cross-section of expected returns," *The Journal of Finance*, vol. 51, no. 1, pp. 3–53, 1996.

[41] J. Lewellen and S. Nagel, "The conditional capm does not explain asset-pricing anomalies," *Journal of Financial Economics*, vol. 82, no. 2, pp. 289–314, 2006.

[42] O. Boguth, M. Carlson, A. Fisher, and M. Simutin, "Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas," *Journal of Financial Economics*, vol. 102, no. 2, pp. 363–389, 2011.

[43] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, no. 6, pp. 066–111, 2004.

[44] B. H. Solnik, "Why not diversify internationally rather than domestically?" *Financial Analysts Journal*, vol. 30, no. 4, pp. 48–54, 1974.

[45] S. L. Lauritzen, "Graphical models," *Oxford Statistical Science Series*, vol. 1999, 1999.

[46] T. Zhao, X. Li, H. Liu, K. Roeder, J. Lafferty, and L. Wasserman, "Huge package for high-dimensional undirected graph estimation," *R Package*, vol. 2015, 2015.

[47] G. Bekaert, R. J. Hodrick, and X. Zhang, "International stock return comovements," *The Journal of Finance*, vol. 64, no. 6, pp. 2591–2626, 2009.

[48] F. Longin and B. Solnik, "Extreme correlation of international equity markets," *The Journal of Finance*, vol. 56, no. 2, pp. 649–676, 2001.

[49] G. Bekaert, C. R. Harvey, and A. Ng, "Market integration and contagion," *Journal of Business*, vol. 78, no. 1, pp. 1–31, 2005.

[50] A. W. Edwards and L. L. Cavalli-Sforza, "A method for cluster analysis," *Biometrics*, vol. 21, no. 2, pp. 362–375, 1965.

[51] R. Killick, K. Haynes, I. Eckley, P. Fearnhead, and J. Lee, "Change point: Method for change point detection," *R package*, vol. 2016, 2016.

[52] J. H. Cochrane, "Presidential address: Discount rates," *The Journal of Finance*, vol. 66, no. 4, pp. 1047–1108, 2011.

[53] C. R. Harvey, Y. Liu, and H. Zhu, "... and the cross-section of expected returns," *Review of Financial Studies*, vol. 29, no. 1, pp. 5–68, 2016.

[54] J. Y. Campbell, M. Lettau, B. G. Malkiel, and Y. Xu, "Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk," *The Journal of Finance*, vol. 56, no. 1, pp. 1–43, 2001.

[55] B. Herskovic, B. Kelly, H. Lustig, and S. Van Nieuwerburgh, "The common factor in idiosyncratic volatility: Quantitative asset pricing implications," *Journal of Financial Economics*, vol. 119, no. 2, pp. 249–283, 2016.

[56] L. Menzly and O. Ozbas, “Market segmenation and cross-predictablility of returns,” *The Journal of Finance*, vol. 64, no. 4, pp. 1555–1580, 2010.

[57] G. Csardi, “Igraph package for network analysis and visualization,” *R package*, vol. 2015, 2015.

[58] J. Chevalier and G. Ellison, “Career concerns of mutual fund managers,” *The Quarterly Journal of Economics*, vol. 114, no. 2, pp. 389–432, 1999.

[59] M. Kacperczyk, C. Sialm, and L. Zheng, “On the industry concentration of actively managed equity mutual funds,” *The Journal of Finance*, vol. 60, no. 4, pp. 1983–2011, 2005.

[60] K. M. Cremers and A. Petajisto, “How active is your fund manager? a new measure that predicts performance,” *The Review of Financial Studies*, vol. 22, no. 9, pp. 3329–3365, 2009.

[61] K. Daniel, M. Grinblatt, S. Titman, and R. Wermers, “Measuring mutual fund performance with characteristic-based benchmarks,” *The Journal of Finance*, vol. 52, no. 3, pp. 1035–1058, 1997.

[62] Y. Amihud and R. Goyenko, “Mutual fund’s r^2 as predictor of performance,” *The Review of Financial Studies*, vol. 26, no. 3, pp. 667–694, 2013.

[63] M. Kacperczyk and A. Seru, “Fund manager use of public information: New evidence on managerial skills,” *The Journal of Finance*, vol. 62, no. 2, pp. 485–528, 2007.

[64] M. Kacperczyk, S. V. Nieuwerburgh, and L. Veldkamp, “Time-varying fund manager skill,” *The Journal of Finance*, vol. 69, no. 4, pp. 1455–1484, 2014.

[65] E. F. Fama and K. R. French, “Luck versus skill in the cross-section of mutual fund returns,” *The journal of finance*, vol. 65, no. 5, pp. 1915–1947, 2010.

[66] J. B. Berk and J. H. Van Binsbergen, “Measuring skill in the mutual fund industry,” *Journal of Financial Economics*, vol. 118, no. 1, pp. 1–20, 2015.

[67] R. B. Evans, “Mutual fund incubation,” *The Journal of Finance*, vol. 65, no. 4, pp. 1581–1611, 2010.

[68] J. Chevalier and G. Ellison, “Risk taking by mutual funds as a response to incentives,” *Journal of Political Economy*, vol. 105, no. 6, pp. 1167–1200, 1997.

[69] E. R. Sirri and P. Tufano, “Costly search and mutual fund flows,” *The journal of finance*, vol. 53, no. 5, pp. 1589–1622, 1998.

[70] G. Hoberg and G. Phillips, "Product market synergies and competition in mergers and acquisitions: A text-based analysis," *The Review of Financial Studies*, vol. 23, no. 10, pp. 3773–3811, 2010.

[71] Y. Y. Tseng and W.-H. Li, "Classification of protein functional surfaces using structural characteristics," *Proceedings of the National Academy of Sciences*, vol. 109, no. 4, pp. 1170–1175, 2012.

[72] G. Hoberg and V. Maksimovic, "Redefining financial constraints: A text-based analysis," *The Review of Financial Studies*, vol. 28, no. 5, pp. 1312–1352, 2014.

[73] G. Hoberg and G. Phillips, "Text-based network industries and endogenous product differentiation," *Journal of Political Economy*, vol. 124, no. 5, pp. 1423–1465, 2016.

[74] A. Koch, "Herd behavior and mutual fund performance," *Management Science*, vol. 63, no. 11, pp. 3849–3873, 2016.

[75] S. J. Brown and W. N. Goetzmann, "Performance persistence," *The Journal of Finance*, vol. 50, no. 2, pp. 679–698, 1995.

[76] M. J. Gruber, "Another puzzle: The growth in actively managed mutual funds," *The Journal of Finance*, vol. 51, no. 3, pp. 783–810, 1996.

[77] H. White, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, vol. 48, pp. 817–838, 1980.

[78] E. F. Fama and J. D. MacBeth, "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, vol. 81, no. 3, pp. 607–636, 1973.

[79] J. Chen, H. Hong, M. Huang, and J. D. Kubik, "Does fund size erode mutual fund performance? the role of liquidity and organization," *The American Economic Review*, vol. 94, no. 5, pp. 1276–1302, 2004.

VITA

Fengrong Wei completed her Ph.D. in Finance at the Scheller College of Business, Georgia Institution of Technology. Fengrong's research interests include fintech, machine learning, empirical asset pricing, international finance, and mutual fund.